CHAOS, CONCENTRATION, AND MULTIPLE VALLEYS

SOURAV CHATTERJEE

ABSTRACT. Disordered systems are an important class of models in statistical mechanics, having the defining characteristic that the energy landscape is a fixed realization of a random field. Examples include various models of glasses and polymers. They also arise in other areas, like fitness models in evolutionary biology. The ground state of a disordered system is the state with minimum energy. The system is said to be chaotic if a small perturbation of the energy landscape causes a drastic shift of the ground state. We present a rigorous theory of chaos in disordered systems that confirms long-standing physics intuition about connections between chaos, anomalous fluctuations of the ground state energy, and the existence of multiple valleys in the energy landscape. Combining these results with mathematical tools like hypercontractivity, we establish the existence of the above phenomena in eigenvectors of GUE matrices, the Kauffman-Levin model of evolutionary biology, directed polymers in random environment, a subclass of the generalized Sherrington-Kirkpatrick model of spin glasses, the discrete Gaussian free field, and continuous Gaussian fields on Euclidean spaces. We also list several open questions.

CONTENTS

1. Introduction

Let us begin with a motivating example. Let $\mathbf{g} = (g_v)_{v \in \mathbb{Z}^2}$ be a collection of i.i.d. standard Gaussian random variables. A $(1+1)$-dimensional directed polymer of length n is a sequence of n adjacent points in $\mathbb{Z}^2$, beginning at the origin, such that each successive point is either to the right or above the previous point. The energy of a polymer $p = (v_0, \ldots, v_{n-1})$ in the Gaussian random environment $\mathbf{g}$ is defined as

$$E(p) := -\sum_{i=0}^{n-1} g_{v_i}.$$

The 'ground state' of the system is the polymer path with minimum energy, which we denote by P. One of the main goals of this paper is to understand a particular feature of the ground state, known as the chaos property. It says, roughly, that a small perturbation of the environment gives rise to a new ground state that is almost disjoint from the original one.

There is a standard way to define a perturbation of a Gaussian environment. If $\mathbf{g}'$ is an independent copy of $\mathbf{g}$, and we define the perturbed environment $\mathbf{g}^t := e^{-t}\mathbf{g} + \sqrt{1 - e^{-2t}}\mathbf{g}'$, then $\mathbf{g}^t$ is again a standard Gaussian random environment. The parameter t is a measure of the amount of perturbation. This definition arises naturally from running an Ornstein-Uhlenbeck flow at each vertex for time t.

Formally, the property of chaos of the ground state means that there exists $t_0(n)$ such that $t_0(n) \to 0$ and $\sup_{t \geq t_0(n)} \mathbb{E}|P \cap P^t| = o(n)$ as $n \to \infty$, where P^t is the minimum energy path in the environment $\mathbf{g}^t$, and $|P \cap P^t|$ is the number of vertices common to the two paths. The definition of 'almost disjoint' in this way makes sense, because P and P^t are both of length n.

Although this is a widely studied phenomenon in the theoretical physics literature on directed polymers (see e.g. [38], [83], [52], [25], [39], [66]), there are no rigorous results. Using the techniques of this paper, we can prove the following theorem.

Theorem 1.1. *Fix n, and let $t_0 = (\log n)^{-1/2}$. Then for all $t \geq t_0$,*

$$\mathbb{E}|P \cap P^t| \leq \frac{Cn}{\sqrt{\log n}},$$

where C is a universal constant.

Of course, this result only proves chaos in principle. The factor of $\sqrt{\log n}$ is too slowly growing to be of any practical significance, even when n is of the order of the Avogadro number.

The key to our approach is a seemingly new connection between the fluctuations of the ground state energy and the stability of the minimum energy path. Suppose τ is an Exponential random variable with mean 1, independent of all else. Then we have the relation

$$\mathrm{Var}(\min_p E(p)) = \mathbb{E}|P \cap P^\tau|.$$

This comes as a consequence of a far more general result, that we are going to describe in the following pages. Given this formula, there are still two tasks remaining: (a) to show that the variance is $o(n)$, and (b) to prove a Tauberian theorem that extracts a bound on $\mathbb{E}|P \cap P^t|$ for fixed t from a bound on $\mathbb{E}|P \cap P^\tau|$. We carry out task (a) in Section 8 and task (b) in Section 3 (more specifically, via Theorem 3.2, which is one of the main results of this paper).

Let us now present our general framework, that encompasses the polymer model as a special case. We treat the energy landscape of the model as a giant Gaussian random vector, and prove general theorems about Gaussian vectors that imply results like Theorem 1.1. There are many other examples that fall into this framework; besides polymers, the ones that have been treated in this paper include spin glass models, random matrices, fitness models of evolutionary biology, the discrete Gaussian free field, and continuous Gaussian fields on Euclidean spaces.

Let S be a finite set and $\mathbf{X} = (X_i)_{i \in S}$ be a centered Gaussian random vector with possibly dependent coordinates. (In the context of the polymer example, think of S as the set of all directed polymers of length n starting at the origin, and $X_i = -E(i)$ for a path $i \in S$.) Let

$$R(i, j) := \mathrm{Cov}(X_i, X_j), \quad \sigma^2 := \max_{i \in S} \mathrm{Var}(X_i).$$

(Again, for polymers $R(i, j) = |i \cap j|$, and $\sigma^2 = n$.) We will often refer to $\mathbf{X}$ as a 'Gaussian field'. The elements of S will be alternately called 'states' or 'indices' or 'sites' or 'coordinates'. The two central objects of interest in this paper are (i) the maximum of the Gaussian field $\mathbf{X}$,

$$M := \max_{i \in S} X_i,$$

and (ii) the location of the maximum,

$$I := \mathrm{argmax}_{i \in S} X_i.$$

To ensure that I is well-defined, we assume the non-degeneracy condition

$$(1) \qquad\qquad \mathbb{P}(X_i \neq X_j) = 1 \ \text{ for each } i \neq j.$$

We study M through its mean and variance

$$m := \mathbb{E}(M), \quad v := \mathrm{Var}(M).$$

Let us alert the reader that we will use the symbols $\mathbf{X}$, M, m, v, $R(i, j)$, and I throughout the paper to mean what they stand for here, often without explicit reference to the above definitions.

The following is a well-known result about the fluctuations of Gaussian maxima.

Proposition 1.2. *Irrespective of the correlation structure of the vector* $\mathbf{X}$, *we always have $v \leq \sigma^2$.*

In words, this means that the order of fluctuations of the maximum cannot be larger than the order of the fluctuations of the most fluctuating coordinate. This inequality was proved by Houdré [36], although the method of proof seems to be implicit in the much earlier work of Nash [55], and the works of Chernoff [18], Chen [17], and Houdré and Kagan [37] on the so-called Poincaré inequality for the Gaussian measure.

There is a famous 'advanced version' of the Poincaré inequality, called the Gaussian isoperimetric inequality, independently invented by Borell [12] and Sudakov and Tsirelson [70], that gives tail bounds instead of simply a variance inequality. A striking consequence of the isoperimetric inequality is the following result of Tsirelson, Ibragimov, and Sudakov [81].

Proposition 1.3. *For any $r \geq 0$,*

$$\mathbb{P}\big(M - m \geq r\big) \leq e^{-r^2/2\sigma^2},$$

and the same bound holds for $P(M - m \leq -r)$ as well.

Although the above result is often called 'Borell's inequality', it is clearly not a fair nomenclature. Since it is too cumbersome to call it the 'Borell-Tsirelson-Ibragimov-Sudakov inequality', we will simply refer to it as Proposition 1.3 in this manuscript.

Note that although Proposition 1.3 is a deep and powerful result, conceptually it does not say a lot more than Proposition 1.2, since it implies, just as Proposition 1.2, that the fluctuations of the maximum can be at most of order σ^2. In particular, it is a crude worst case bound that does not use the correlation structure of $\mathbf{X}$. However, this is all that one can obtain from the classical theory of concentration of measure (see e.g. Ledoux [47]).

Here is where our investigation begins. What happens if v is very small compared to σ^2? As we will see, this is in fact the rule rather than the exception in interesting examples. The main point of this paper is that the condition $v \ll \sigma^2$ ushers in a whole host of interesting structure on the field $\mathbf{X}$. Indeed, the structure is so interesting and pervasive that the condition seems to deserve a name of its own. When it happens, we will say that the Gaussian field $\mathbf{X}$ exhibits 'superconcentration'. The notion can be precisely defined only in terms of a sequence of Gaussian fields rather than a single one. Accordingly, let $(\mathbf{X}_n)_{n \geq 1}$ be a sequence of centered Gaussian fields, where $\mathbf{X}_n$ is defined on a finite set S_n. Let

$$\sigma_n^2 := \max_{i \in S_n} \mathrm{Var}(X_{n,i}), \quad M_n := \max_{i \in S_n} X_{n,i}, \quad m_n := \mathbb{E}(M_n), \quad v_n = \mathrm{Var}(M_n).$$

Definition 1.4. *We say that the sequence of Gaussian fields $(\mathbf{X}_n)_{n \geq 1}$ 'has superconcentrated maximum' or simply 'is superconcentrated' if $v_n = o(\sigma_n^2)$ as $n \to \infty$.*

Here, as usual, $a_n = o(b_n)$ means that $\lim_{n \to \infty} a_n/b_n = 0$. In practice, we will simply say that $\mathbf{X}$ is superconcentrated if it is implicitly the nth member of a sequence of fields having superconcentrated maximum. We will

give many examples of superconcentrated Gaussian fields in the subsequent sections to demonstrate the 'rule rather than exception' claim.

Incidentally, physicists often refer to the superconcentration phenomenon as 'anomalous fluctuations' (see e.g. [39]). However, it is not a well-defined notion (in particular, they don't connect it with classical concentration, and 'anomalous fluctuations' can also mean larger fluctuations than usual); we feel that our terminology is more evocative and precise.

Let us now describe some of the consequences of superconcentration. A summary of the results is contained in Theorem 1.8, but we first need to define some concepts.

An important property of Gaussian fields that has been studied in various special examples by physicists but has almost no presence in rigorous mathematics, is the property of chaos. We have already had a discussion of this in the context of polymers, so let us now make a general definition. Let $\mathbf{X}'$ be an independent copy of $\mathbf{X}$. For each $t \in [0, \infty)$, let

$$\mathbf{X}^t := e^{-t}\mathbf{X} + \sqrt{1 - e^{-2t}}\mathbf{X}'.$$

Note that $\mathbf{X}^t$ has the same distribution as $\mathbf{X}$, that is, the transformation $\mathbf{X} \to \mathbf{X}^t$ is a distribution preserving perturbation of $\mathbf{X}$. We will be mostly interested in small perturbations, i.e., small t. As mentioned before, this is a natural way to define perturbations of Gaussian fields because of its intimate relation to Ornstein-Uhlenbeck diffusions. Let I^t be the state at which the maximum is attained in $\mathbf{X}^t$, that is,

$$I^t := \operatorname{argmax}_{i \in S} X_i^t.$$

Note that I^t is well-defined by assumption (1), and that $I^0 = I$. We will say that the field $\mathbf{X}$ is 'chaotic' if I^t is highly unstable, that is a small change in the value of t causes, with high probability, a drastic change in I^t. There are, of course, various notions to be made precise here. First of all, what is meant by a drastic change in I^t? As we will see, in most examples two states $i, j \in S$ are 'drastically different' if $R(i, j)$ is very small, typically $R(i, j) \ll \sigma^2$. Thus, we may formulate chaos for I^t to mean that for $t = o(1)$, $\mathbb{E}(R(I^0, I^t)) \ll \sigma^2$. Secondly, what is a 'small change in t'? This we will take at face value, i.e. small means small, in relation to nothing else.

However, none of this is precise. As before, the only way to make a completely meaningful definition is via sequences.

Accordingly, let $(\mathbf{X}_n)_{n \geq 1}$ be a sequence of Gaussian fields as in Definition 1.4. Let $\mathbf{X}'_n$ be an independent copy of $\mathbf{X}_n$ and define, as above, the perturbed fields

$$\mathbf{X}_n^t := e^{-t}\mathbf{X}_n + \sqrt{1 - e^{-2t}}\mathbf{X}'_n.$$

Again, as above, let $I_n^t = \operatorname{argmax}_{i \in S_n} X_{n,i}^t$. Let R_n be the covariance kernel of $\mathbf{X}_n$. We are now ready to give a precise definition of the chaos property.

Definition 1.5. *We say that the location of the maximum in $\mathbf{X}_n$ exhibits 'chaos' or simply that $\mathbf{X}_n$ exhibits chaos if there is a sequence $t_n \to 0$ such that $\mathbb{E}R_n(I_n^0, I_n^{t_n}) = o(\sigma_n^2)$ as $n \to \infty$.*

Note that here we defined chaos in terms of the decay of $\mathbb{E}R_n(I_n^0, I_n^{t_n})$, instead of $\sup_{t \geq t_n} \mathbb{E}R_n(I_n^0, I_n^t)$ as we did for the polymer example. It turns out (Theorem 3.2) that $\mathbb{E}R_n(I_n^0, I_n^t)$ is always a decreasing function of t, and therefore the two definitions are equivalent.

Next let us turn our attention to the so-called 'multiple valley picture'. Often, we have the situation that a Gaussian field $\mathbf{X}$ has many 'drastically different' sites at which the global maximum is nearly achieved. It is called 'multiple valleys' instead of 'multiple peaks' because the physicists like to put a minus sign. We will, however, call it the multiple peaks phenomenon to avoid any confusion. As before, we attempt to give a precise definition via sequences of fields. All notation is the same as before.

Definition 1.6. *We say that $\mathbf{X}_n$ has multiple peaks (MP) if there exist $l_n \to \infty$, $\epsilon_n = o(\sigma_n^2)$, $\delta_n = o(m_n)$ and $\gamma_n \to 0$ such that with probability at least $1 - \gamma_n$, there is a set $A \subseteq S_n$ satisfying*

    (a) *$|A| = l_n$,*
    (b) *$R_n(i, j) < \epsilon_n$ for each $i, j \in A$, $i \neq j$, and*
    (c) *$X_i \geq M_n - \delta_n$ for each $i \in A$.*

Note that the condition $\delta_n = o(m_n)$ is natural, because X_i is nearly maximum if $X_i = M_n - o(M_n)$. However, this is not the form that is conjectured in the physics models. Indeed, since the fluctuations of M_n are of order $\sqrt{v_n}$, the physicists seem to think that multiple peaks, in the above sense, should occur whenever $\delta_n \gg \sqrt{v_n}$, or at least when $\delta_n = o(\sigma_n)$. This leads us to the definition of a stronger notion of multiple peaks.

Definition 1.7. *We say that $\mathbf{X}_n$ has strong multiple peaks if the multiple peaks condition is satisfied with $\delta_n = o(\sigma_n)$ instead of $\delta_n = o(m_n)$.*

The first main result of this paper, stated below, shows that the properties of superconcentration, chaos, multiple peaks, and strong multiple peaks are all intimately related to each other. This may not be surprising from a physicist's point of view, but this is the first time that these widely observed phenomena have been formulated and connected by rigorous mathematics.

Theorem 1.8. *For any sequence of Gaussian fields $(\mathbf{X}_n)_{n \geq 1}$ satisfying the non-degeneracy condition (1) we have*

$$\text{Strong MP} \; \overset{\Longrightarrow}{\underset{\nLeftarrow}{}} \; \text{Superconcentration} \iff \text{Chaos}.$$

Moreover, under the 'positivity assumption' that $R_n(i, j) \geq 0$ for each n and $i, j \in S_n$, we have the more complete picture:

$$\text{Strong MP} \; \overset{\Longrightarrow}{\underset{\nLeftarrow}{}} \; \text{Superconcentration} \iff \text{Chaos} \; \overset{\Longrightarrow}{\underset{\nLeftarrow}{}} \; \text{MP}.$$

The counterexample that shows chaos $\not\Rightarrow$ strong MP is particularly surprising and goes against common intuition. It shows that, contrary to what one may think, chaos *is not necessarily caused* by the existence of multiple peaks.

Theorem 1.8 is stated as a limiting result, but we do have precise quantitative bounds for all parts of the theorem. These will be presented in Section 3, where we give the proof of Theorem 1.8. One result from Section 3 is worth mentioning here, since it provides the foundation for all subsequent work by connecting the value of the maximum with the location of the maximum in a Gaussian field. It is Lemma 3.1, which says that if τ is an Exponential random variable with mean 1, independent of all else, then we have the formula

$$v = \mathbb{E}(R(I^0, I^\tau)).$$

We have already stated the version of this formula for polymers. The proof of this identity would be easy for any expert on the Ornstein-Uhlenbeck semigroup — indeed, it is implicit in the classical proofs of Propositions 1.2 and 1.3 — but the author has not seen it explicitly written down anywhere in the published literature. Interestingly, it was brought to our notice that the identity does make an appearance (in a slightly different form) in a recent manuscript of Nourdin and Viens [58] that was prepared at the same time as the first version of this paper was being written.

In spite of its cuteness the identity is not very useful on its own, since we are interested in $\mathbb{E}(R(I^0, I^t))$ for fixed t. This question is handled in Theorem 3.2, where we apply a Tauberian argument to the above representation of v to show that for each t,

$$0 \le \mathbb{E}(R(I^0, I^t)) \le \frac{v}{1 - e^{-t}}, \quad \text{and}$$
$$v \le \sigma^2(1 - e^{-t}) + \mathbb{E}(R(I^0, I^t))e^{-t}.$$

It turns out that these upper and lower bounds suffice to show the equivalence of superconcentration and chaos (we show this in Section 3).

Before presenting further results, let us discuss some of the literature. As we mentioned before, the phenomena of superconcentration, chaos, and multiple peaks have not been systematically studied in the mathematics literature, so there are essentially very few references. The closest thing to chaos in the domain of rigorous mathematics is the notion of noise-sensitivity, although that refers mainly to correlations between functions. The literature on noise-sensitivity in computer science is sizable; it mostly involves sensitivity of scalar functions of Boolean random variables to random noise, which is not so relevant to us. In the probability world, a very notable paper on the subject is due to Benjamini, Kalai, and Schramm [6]. A subsequent paper [7] by the same authors is more relevant for what we do in this article.

One truly significant contribution to what we call superconcentration is due to Talagrand ([71], Theorem 1.5), which has been the source of many applications (including [7]). Talagrand's result provides a way to improve

variance bounds like Proposition 1.2 under certain situations by a 'factor of $\log n$'. Talagrand's breakthrough idea was to use the tool of hypercontractivity (discovered by Nelson [57]) to improve variance bounds. We will use this method, in conjunction with ideas from Benjamini-Kalai-Schramm [7] and our Theorem 1.8, to prove the existence of chaos in directed polymers in Section 8.

Another contribution of Talagrand in our context is the proof of strong MP $\Rightarrow$ superconcentration, which follows essentially from a sketch at the end of Section 8.3 in his landmark paper [72]. Unfortunately, the author has not yet been able to find a use for this remarkable implication, since proving strong MP seems to be always more difficult than proving any of the other phenomena.

In the physics literature, there is a long, folklorish history of studying the phenomena of superconcentration (via 'fluctuation exponents'), chaos, and multiple valleys. For instance, the implication that superconcentration $\Rightarrow$ chaos is the central theme of Fisher and Huse [25], who investigated it in the context of directed polymers in a random environment. Examples of other highly cited works in this area are those of McKay, Berger, and Kirkpatrick [50], Huse, Henley, and Fisher [38], Bray and Moore [13], Zhang [83] and Mézard [52].

Let us now return to the discussion of our results. As we mentioned above, the only rigorous tool available at present that can establish superconcentration is Talagrand's method of using hypercontractivity to improve variance bounds. Although this works in many situations (including some of our examples in this paper), it gives only a '$\log n$ correction', and usually does not suffice to break the barrier of 'improvements in powers of n'.

One of the main goals of this work is to break this wall by finding an alternative technique. We have only had partial success in this direction, but what we have may lead to further progress. Under a certain condition that we call 'extremality', we are able to get improvements in powers of n in highly nontrivial models like certain cases of the generalized Sherrington-Kirkpatrick model of spin glasses. The success is 'partial' because, for instance, we are not able to cover the original 2-spin SK model.

The notion of extremality of a Gaussian field like $\mathbf{X}$ is defined as follows. It can be shown (see Lemma 2.1) that irrespective of the correlation structure, we have $m \leq \sqrt{2\sigma^2 \log |S|}$, with near equality if the coordinates are i.i.d. $N(0, \sigma^2)$. We say that the field $\mathbf{X}$ is 'extremal' if $m \simeq \sqrt{2\sigma^2 \log |S|}$. Of course this makes sense only if we consider sequences.

Definition 1.9. *We say that $(\mathbf{X}_n)_{n \geq 1}$ is extremal if*

$$\lim_{n \to \infty} \frac{m_n}{\sqrt{2\sigma_n^2 \log |S_n|}} = 1.$$

Note that although extremality may seem to indicate that the coordinates of $\mathbf{X}_n$ are approximately independent for large n, that is not true. Extremality can hold even in models with high degrees of dependence between sites, like

the Gaussian free field (proved by Bolthausen, Deuschel, and Giacomin [11]), branching random walks (proved by Biggins [9]), and some nontrivial spin glasses (treated in Section 9). The second main result of this paper, proved in Section 5, is the following.

Theorem 1.10. *For any sequence of Gaussian fields $(\mathbf{X}_n)_{n \geq 1}$ with state spaces S_n growing in size to infinity, we have*

$$Extremality \implies Chaos.$$

Let us now present one example of a concrete variance bound that can be used to establish superconcentration. The following theorem is proved in Section 5, where it is deduced from a quantitative version of Theorem 1.10. We shall use this result to establish the presence of chaos in certain models of spin glasses in Section 9. Moreover, the variance bound given by the following theorem, wherever it applies, gives 'corrections in powers of n' instead of log corrections as given by hypercontractivity.

Theorem 1.11. *Consider the field $\mathbf{X} = (X_i)_{i \in S}$. Suppose $R(i,i) = \sigma^2$ for all i. For each $i, j \in S$, let $r_{ij} := R(i,j)/\sigma^2$. Let*

$$\beta := \left(\frac{\log\log|S| + \log \sum_{i,j \in S} |S|^{-2/(1+r_{ij})}}{\log|S|} \right)^{1/4}.$$

Then $v \leq C\sigma^2\beta$ and for any $t \geq 0$, $\mathbb{E}(R(I^0, I^t)) \leq \frac{C\sigma^2\beta}{\sqrt{1-e^{-2t}}}$, where C is a universal constant.

The bound can be interpreted easily by considering i.i.d. standard Gaussians. If $r_{ij} = 0$ for all $i \neq j$ and $r_{ii} = 1$ for all i, we have $\sum_{i,j \in S} |S|^{-2/(1+r_{ij})} \leq 2$, which proves superconcentration (although not the correct order bound on the variance in this case). In general, Theorem 1.11 gives a way of proving superconcentration and chaos when 'most correlations are small', but may not work in many situations.

The rest of the paper is organized as follows. In Section 2, we state some well-known results about Gaussian random variables that will be useful for us later on. In Section 3, we prove Theorem 1.8, together with a number of results that give quantitative versions of the various implications in Theorem 1.8. In Section 4, we give a brief introduction to the concept of hypercontractivity for the Ornstein-Uhlenbeck semigroup and how to use it for proving superconcentration. In Section 5 we prove Theorems 1.10 and 1.11. Finally, in Sections 6 through 11, we work out a number of examples. This includes applications to eigenvectors of random matrices (Section 6), the Kauffman-Levin NK fitness model of evolutionary biology (Section 7), directed polymers in random environment (Section 8), the generalized Sherrington-Kirkpatrick model of spin glasses (Section 9), the discrete Gaussian free field (Section 10), and Gaussian fields on Euclidean spaces (Section 11).

Let us now mention some conventions that we will follow in this paper. First of all, we must declare that the constant C is going to stand for any generic universal constant, whose value may change from line to line. This is an invaluable help in lightening notation. We will generally denote scalar variables and elements of $\mathbb{R}^2$ by ordinary italic font variable names like x, y, u, etc. We will use boldface in dimensions higher than 2. Finally, let us reiterate that the symbols $\mathbf{X}, \mathbf{X}^t, M, m, v, \sigma^2, I^t$, and $R(i, j)$ will be used without reference to denote what they denote in this section.

2. SOME BASIC FACTS ABOUT GAUSSIAN RANDOM VARIABLES

In this section, as elsewhere, we continue to use the notation defined in Section 1. In particular, $\mathbf{X}$, $\mathbf{X}^t$, m, v, σ^2, $R(i, j)$, and I^t stand for the same objects as before. We state some very well-known facts about Gaussian random variables and vectors that will be of repeated use for us in the rest of the manuscript. Two such facts, namely Proposition 1.2 and Proposition 1.3, have already been stated in Section 1.

Size of the maximum. Just like the variance, the expected value of the maximum of a Gaussian field also has a general, worst case bound. The bound is much easier to establish than the variance bound, so we give the proof right here.

Lemma 2.1. *We have the general bound*

$$m \leq \sqrt{2\sigma^2 \log |S|}.$$

Moreover if $|S| \geq 2$ then for any $p \geq 1$,

$$\mathbb{E}|M|^p \leq \mathbb{E} \max_i |X_i|^p \leq C(p)\sigma^p (\log |S|)^{p/2},$$

where $C(p)$ is a constant that depends only on p.

Remark. We do not need that $(X_i)_{i \in S}$ are Gaussian for the bound on the expectation; the proof goes through for any collection random variables with Gaussian tails, irrespective of the dependence among them. This observation will be used a few times in the sequel.

Proof. Without loss of generality, assume that $\sigma^2 = 1$. Then for any $\beta > 0$,

$$m = \frac{1}{\beta}\mathbb{E}(\log e^{\beta M})$$

$$\leq \frac{1}{\beta}\mathbb{E}\left(\log \sum_{i \in S} e^{\beta X_i}\right)$$

$$\leq \frac{1}{\beta}\log \sum_{i \in S} \mathbb{E}(e^{\beta X_i}) \leq \frac{\beta}{2} + \frac{\log |S|}{\beta}.$$

Optimizing over β, we establish the first claim. For the second, one just has to combine the bound on m with Proposition 1.3, and observe that $\max |X_i|$ is the maximum of the concatenation of the vectors $\mathbf{X}$ and $-\mathbf{X}$. $\qquad\square$

Slepian's lemma and Sudakov minoration. The following result is an indispensable tool in the study of Gaussian processes. It was discovered by Slepian [68] and goes by the name of 'Slepian's lemma'.

Lemma 2.2. *Suppose* $\mathbf{X} = (X_i)_{i \in S}$ *and* $\mathbf{Y} = (Y_i)_{i \in S}$ *are centered Gaussian random vectors with* $\mathbb{E}(X_i^2) = \mathbb{E}(Y_i^2)$ *for each* i *and* $\mathbb{E}(X_i X_j) \geq \mathbb{E}(Y_i Y_j)$ *for each* i, j. *Then for each* $x \in \mathbb{R}$,

$$\mathbb{P}(\max_i X_i > x) \leq \mathbb{P}(\max_i Y_i > x).$$

In particular, $\mathbb{E}(\max_i X_i) \leq \mathbb{E}(\max_i Y_i)$.

The next result is a close analog of Slepian's inequality, known as the Sudakov minoration lemma. For a proof, see Lemma 2.1.2 in [75].

Lemma 2.3. *Suppose* a *is a constant such that* $\mathbb{E}(X_i - X_j)^2 \geq a$ *for all* $i, j \in S$, $i \neq j$. *Then*

$$m \geq Ca\sqrt{\log |S|},$$

where C *is a positive universal constant.*

Mills ratio bounds. The following pair of inequalities is collectively known as the Mills ratio bounds. For a standard Gaussian random variable Z, for any $x > 0$, we have

$$\frac{xe^{-x^2/2}}{\sqrt{2\pi}(1+x^2)} \leq \mathbb{P}(Z > x) \leq \frac{e^{-x^2/2}}{x\sqrt{2\pi}}.$$

The proof is not difficult, and may be found in numerous standard texts on probability and statistics. The inequalities in the above form were probably first proven by Gordon [29]. Along similar lines, one can also prove the inequality

$$\mathbb{P}(Z > x) \leq e^{-x^2/2}, \tag{2}$$

which follows simply by optimizing over $\mathbb{P}(Z > x) \leq e^{-\theta x}\mathbb{E}(e^{\theta Z})$.

Gaussian integration by parts. Suppose Z is a standard Gaussian random variable, and $f : \mathbb{R} \to \mathbb{R}$ is an absolutely continuous function. If $\mathbb{E}|f'(Z)| < \infty$, then one can argue that

$$\mathbb{E}|Zf(Z)| \leq C\mathbb{E}|f'(Z)| + C < \infty, \tag{3}$$

where C is a universal constant. Moreover, a standard application of integration by parts gives the well-known identity

$$\mathbb{E}(Zf(Z)) = \mathbb{E}(f'(Z)).$$

This identity can be easily generalized to a Gaussian random vector like $\mathbf{X}$, as follows. If $f : \mathbb{R}^S \to \mathbb{R}$ is an absolutely continuous function such that $\|\nabla f(\mathbf{X})\|$ has finite expectation, then for any $i \in S$,

$$\mathbb{E}(X_i f(\mathbf{X})) = \sum_{j \in S} R(i,j)\mathbb{E}(\partial_j f(\mathbf{X})),$$

where $\partial_j f$ denotes the partial derivative of f along the jth coordinate. This identity can be derived from the previous one simply by writing $\mathbf{X}$ as a linear transformation of a vector of i.i.d. standard Gaussian random variables. The author encountered this useful version of the integration-by-parts identity in [74], Appendix A.6.

3. Structure of a superconcentrated Gaussian field

The goal of this section is to prove Theorem 1.8. Throughout this section, as everywhere else, we will freely use notation from Section 1 (like $\mathbf{X}^t$, $R(i,j)$, m, v, I^t, and M) without explicit reference. We will divide the proof into a number of subsections, one devoted to each part of the proof. The theorems of this section are all interesting in their own right, because they give quantitative versions of the various implications of Theorem 1.8.

3.1. **Superconcentration is equivalent to Chaos.** We begin with the exact formula for v stated in Section 1.

Lemma 3.1. *Let τ be a standard exponential random variable, independent of everything else. Then we have*

$$(4) \qquad v = \mathbb{E}(R(I^0, I^\tau)).$$

Note that by the Cauchy-Schwarz inequality, this identity implies Proposition 1.2.

Lemma 3.1, combined with an elementary Tauberian argument, leads to the following Theorem, which establishes the equivalence of superconcentration and chaos.

Theorem 3.2. *For each t we have*

$$0 \leq \mathbb{E}(R(I^0, I^t)) \leq \frac{v}{1 - e^{-t}}, \quad and$$
$$v \leq \sigma^2(1 - e^{-t}) + \mathbb{E}(R(I^0, I^t))e^{-t}.$$

Moreover, $\mathbb{E}(R(I^0, I^t))$ is a decreasing function of t.

To see how the bounds imply the claim of equivalence of superconcentration and chaos, consider the following. If, in the notation of Section 1, we have $v_n = o(\sigma_n^2)$, then by choosing $t_n = \sqrt{v_n/\sigma_n^2} = o(1)$ we can guarantee by the first bound that

$$\mathbb{E}(R(I_n^0, I_n^{t_n})) \leq O(\sqrt{v_n}\sigma_n) = o(\sigma_n^2).$$

Again, if we can find $t_n = o(1)$ such that $\mathbb{E}(R(I_n^0, I_n^{t_n})) = o(\sigma_n^2)$, then the second bound shows that $v_n = o(\sigma_n^2)$.

The proof of Lemma 3.1 is done in three simple steps.

Lemma 3.3. *Let $\mathbf{Y} = (Y_i)_{i \in S}$ be a vector of independent standard Gaussian random variables, and let $\mathbf{Y}'$ be an independent copy of $\mathbf{Y}$. Let f :*

$\mathbb{R}^S \to \mathbb{R}$ *be an absolutely continuous with gradient* ∇f *and suppose that* $\mathbb{E}\|\nabla f(\mathbf{Y})\|^2 < \infty$. *For each* $t \in [0, \infty)$, *let* $\mathbf{Y}^t = e^{-t}\mathbf{Y} + \sqrt{1 - e^{-2t}}\mathbf{Y}'$. *Then*

$$(5) \qquad \mathrm{Var}(f(\mathbf{Y})) = \int_0^\infty e^{-t}\mathbb{E}\langle \nabla f(\mathbf{Y}), \nabla f(\mathbf{Y}^t)\rangle \, dt,$$

where $\langle \cdot, \cdot \rangle$ *denotes the usual inner product.*

Proof. Note that

$$\mathrm{Var}(f(\mathbf{Y})) = \mathbb{E}(f(\mathbf{Y})(f(\mathbf{Y}) - f(\mathbf{Y}')))$$

$$= \mathbb{E}\left(-\int_0^\infty f(\mathbf{Y})\frac{d}{dt}f(e^{-t}\mathbf{Y} + \sqrt{1 - e^{-2t}}\mathbf{Y}')dt\right)$$

$$= \mathbb{E}\left(-\int_0^\infty f(\mathbf{Y})\sum_{i \in S}\left(-e^{-t}Y_i + \frac{e^{-2t}Y_i'}{\sqrt{1 - e^{-2t}}}\right)\partial_i f(e^{-t}\mathbf{Y} + \sqrt{1 - e^{-2t}}\mathbf{Y}')dt\right).$$

Now fix $t \in [0, \infty)$, and let

$$\mathbf{V}^t := \sqrt{1 - e^{-2t}}\mathbf{Y} - e^{-t}\mathbf{Y}'.$$

Then $\mathbf{Y}^t$ and $\mathbf{V}^t$ are independent standard Gaussian random vectors and

$$\mathbf{Y} = e^{-t}\mathbf{Y}^t + \sqrt{1 - e^{-2t}}\mathbf{V}^t.$$

Taking any i, and using Gaussian integration-by-parts as outlined in Section 2 (in going from the second to the third line below), we get

$$\mathbb{E}\left(f(\mathbf{Y})\left(e^{-t}Y_i - \frac{e^{-2t}Y_i'}{\sqrt{1 - e^{-2t}}}\right)\partial_i f(e^{-t}\mathbf{Y} + \sqrt{1 - e^{-2t}}\mathbf{Y}')\right)$$

$$= \frac{e^{-t}}{\sqrt{1 - e^{-2t}}}\mathbb{E}\left(f(e^{-t}\mathbf{Y}^t + \sqrt{1 - e^{-2t}}\mathbf{V}^t)V_i^t\partial_i f(\mathbf{Y}^t)\right)$$

$$= e^{-t}\mathbb{E}\left(\partial_i f(\mathbf{Y})\partial_i f(\mathbf{Y}^t)\right).$$

(Note that the condition $\mathbb{E}\|\nabla f(\mathbf{Y})\|^2 < \infty$ and the bound (3) allows us to integrate by parts and interchange integrals and expectations.) This completes the proof of the Lemma. $\square$

Lemma 3.4. *Let* f *be as in Lemma* 3.3. *Then*

$$\mathrm{Var}(f(\mathbf{X})) = \int_0^\infty e^{-t}\sum_{i,j=1}^n R(i,j)\mathbb{E}(\partial_i f(\mathbf{X})\partial_j f(\mathbf{X}^t)) \, dt.$$

Proof. If $R = BB^T$ for some matrix B, then we can assume $\mathbf{X} = B\mathbf{Y}$ where $\mathbf{Y}$ is a standard Gaussian r.v. As before, let $\mathbf{Y}'$ be an independent copy of $\mathbf{Y}$ so that $\mathbf{X}' = B\mathbf{Y}'$ is an independent copy of $\mathbf{X}$. Putting $g(\mathbf{y}) := f(B\mathbf{y})$, and using Lemma 3.3 with g instead of f, an easy computation gives

$$(6) \qquad \sum_{i=1}^n \partial_i g(\mathbf{Y})\partial_i g(\mathbf{Y}^t) = \sum_{i,j=1}^n R(i,j)\partial_i f(\mathbf{X})\partial_j f(\mathbf{X}^t).$$

By Lemma 3.3, this completes the proof. $\square$

Proof of Lemma 3.1. Consider the function $f(\mathbf{x}) := \max_{i \in S} x_i$. We have

$$\partial_i f(\mathbf{x}) = 1_{\{x_i \geq x_j \forall j\}} \quad \text{a.e.}$$

The proof now follows easily from Lemma 3.4. $\hfill\square$

The proof of Theorem 3.2 requires one more lemma. The current proof of the following result is a major simplification (thanks to Michel Ledoux) of the author's proof in the first draft.

Lemma 3.5. *Let* $\mathbf{Y}$ *and* $\mathbf{Y}^t$ *be as in Lemma* 3.3. *Then for any function* $h : \mathbb{R}^n \to \mathbb{R}$ *such that* $\mathbb{E}(h(\mathbf{Y})^2) < \infty$, $\mathbb{E}(h(\mathbf{Y})h(\mathbf{Y}^t))$ *is a nonnegative, non-increasing function of* t.

Proof. Fix $t \geq 0$. Let $\mathbf{Y}''$ be another independent copy of $\mathbf{Y}$. Let

$$\mathbf{Y}^{-t/2} := e^{-t/2}\mathbf{Y} + \sqrt{1 - e^{-t}}\mathbf{Y}''.$$

It is easy to verify (by checking covariances) that the pair $(\mathbf{Y}, \mathbf{Y}^t)$ has the same distribution as the pair $(\mathbf{Y}^{-t/2}, \mathbf{Y}^{t/2})$. Again, it is trivial to see that given $\mathbf{Y}$, the vectors $\mathbf{Y}^{-t/2}$ and $\mathbf{Y}^{t/2}$ are independent and identically distributed. Thus,

$$(7) \qquad \begin{aligned} \mathbb{E}(h(\mathbf{Y})h(\mathbf{Y}^t)) &= \mathbb{E}(h(\mathbf{Y}^{-t/2})h(\mathbf{Y}^{t/2})) \\ &= \mathbb{E}\big((P_{t/2}h(\mathbf{Y}))^2\big), \end{aligned}$$

where

$$P_s h(\mathbf{Y}) := \mathbb{E}(h(\mathbf{Y}^s) \mid \mathbf{Y}).$$

This shows that $\mathbb{E}(h(\mathbf{Y})h(\mathbf{Y}^t))$ is nonnegative. Now, it is easy to verify that $(P_s)_{s \geq 0}$ is a semigroup of operators, that is, $P_s P_t = P_{s+t}$. Using this, note that for any $s, t \geq 0$,

$$\begin{aligned} \mathbb{E}\big((P_t h(\mathbf{Y}))^2 - (P_{t+s}h(\mathbf{Y}))^2\big) &= \mathbb{E}\big((P_t h(\mathbf{Y}^s))^2 - (\mathbb{E}(P_t h(\mathbf{Y}^s) \mid \mathbf{Y}))^2\big) \\ &= \mathbb{E}\big(\text{Var}(P_t h(\mathbf{Y}^s) \mid \mathbf{Y})\big) \geq 0. \end{aligned}$$

Combined with the representation (7), this shows that $\mathbb{E}(h(\mathbf{Y})h(\mathbf{Y}^t))$ is non-increasing in t. $\hfill\square$

Proof of Theorem 3.2. By Lemma 3.5 and the representation (6), we see that $\mathbb{E}(R(I^0, I^t))$ is nonnegative and non-increasing as a function of t. Combining this with Lemma 3.1, we get that for any t,

$$\begin{aligned} v &= \int_0^\infty e^{-s}\mathbb{E}(R(I^0, I^s))ds \\ &\geq \int_0^t e^{-s}\mathbb{E}(R(I^0, I^s))ds \\ &\geq \int_0^t e^{-s}\mathbb{E}(R(I^0, I^t))ds = (1 - e^{-t})\mathbb{E}(R(I^0, I^t)). \end{aligned}$$

Similarly,

$$v \le \sigma^2 \int_0^t e^{-s} ds + \mathbb{E}(R(I^0, I^t)) \int_t^\infty e^{-s} ds$$
$$= \sigma^2(1 - e^{-t}) + \mathbb{E}(R(I^0, I^t))e^{-t}.$$

This completes the proof. □

3.2. **Strong Multiple Peaks implies Superconcentration.** As usual, we give a quantitative version of this result. The difference with other parts of the proof is that this proof is not an original idea of the author; it follows from a sketch given at the end of Section 8.3 in Talagrand's famous treatise on concentration inequalities [72]. To the best of our knowledge, this sketch has never been formulated as a concrete theorem before.

Theorem 3.6. *Suppose l is a positive integer, and $\epsilon > 0$, $\delta > 0$, and $\gamma \in [0, 1]$ are numbers such that with probability $1 - \gamma$, there exists a set $A \subseteq S$ such that*

(a) $|A| \ge l$,
(b) $R(i, j) < \epsilon$ for all $i, j \in A$, $i \ne j$, and
(c) $X_i \ge M - \delta$ for all $i \in A$.

Then $v \le 3\delta^2 + 3l^{-1}\sigma^2 + 3\epsilon + 16\sigma^2 \sqrt{\gamma}$.

It is easy to see from the above result how we get superconcentration if we have a sequence of Gaussian fields as in Section 1 satisfying the strong MP condition with $l_n \to \infty$, $\epsilon_n = o(\sigma_n^2)$, $\delta_n = o(\sigma_n)$ and $\gamma_n \to 0$.

Proof. Let

$$U := \{(i_1, i_2, \ldots, i_l) : i_1, \ldots, i_l \in S, \ R(i_p, i_q) < \epsilon \text{ for all } p \ne q\}.$$

For each $(i_1, \ldots, i_l) \in U$, let

$$Z_{(i_1, \ldots, i_l)} := \frac{1}{l} \sum_{p=1}^l X_{i_p}.$$

A simple computation shows that

$$\operatorname{Var}(Z_{(i_1, \ldots, i_l)}) \le l^{-1}\sigma^2 + \epsilon.$$

Thus, if we let

$$\overline{M} := \max_{(i_1, \ldots, i_l) \in U} Z_{(i_1, \ldots, i_l)},$$

then by Proposition 1.2

(8) $$\operatorname{Var}(\overline{M}) \le l^{-1}\sigma^2 + \epsilon.$$

Now, let $\mathbf{X}'$ be an independent copy of $\mathbf{X}$, and define M', $Z'_{(i_1, \ldots, i_l)}$, and $\overline{M}'$ accordingly. Let E denote the event of the existence of a set A satisfying (a), (b), and (c) for the vector $\mathbf{X}$ and a set A' satisfying (a), (b), and (c) for the vector $\mathbf{X}'$. If E happens, then $|M - \overline{M}| \le \delta$ and $|M' - \overline{M}'| \le \delta$. By the

inequality $(x+y+z)^2 \leq 3(x^2+y^2+z^2)$ and the Cauchy-Schwarz inequality in the third step below, we have

$$
\begin{aligned}
2\mathrm{Var}(M) &= \mathbb{E}(M - M')^2 \\
&= \mathbb{E}((M - M')^2; E) + \mathbb{E}((M - M')^2; E^c) \\
&\leq 6\mathbb{E}((M - \overline{M})^2; E) + 3\mathbb{E}((\overline{M} - \overline{M}')^2; E) \\
&\quad + \left[\mathbb{P}(E^c)\mathbb{E}(M - M')^4\right]^{1/2} \\
&\leq 6\delta^2 + 3\mathbb{E}(\overline{M} - \overline{M}')^2 + (2\gamma\,\mathbb{E}(M - M')^4)^{1/2}.
\end{aligned}
$$

By (8), the second term is bounded by $6(l^{-1}\sigma^2 + \epsilon)$. By Proposition 1.3,

$$
\mathbb{E}(M - M')^4 \leq 16\sigma^4 \int_0^\infty x^3 e^{-x^2/8}dx = 512\sigma^4.
$$

This completes the proof.   □

3.3. Under positivity, Chaos implies Multiple Peaks.

The goal of this subsection is to prove that if $R(i,j) \geq 0$ for all i,j, then the property of chaos guarantees the multiple peaks condition. As before, we have a quantitative statement.

Theorem 3.7. *Suppose $R(i,j) \geq 0$ for all i,j. Then for any integer $l \geq 2$, any $\epsilon \in (0, \sigma^2)$, and any $\delta \in (0, m)$, we have that with probability at least*

$$
\text{(9)} \qquad 1 - 4\sqrt{\frac{vml^3 \log l}{\delta\epsilon}} - 4\left(\frac{v\sigma^4 l^5 \log l}{\delta^3 m\epsilon}\right)^{1/4}
$$

there exists $A \subseteq S$ such that

(a) $|A| = l$,
(b) $R(i,j) < \epsilon$ for all $i,j \in A$, $i \neq j$, and
(c) $X_i \geq M - \delta$ for all $i \in A$.

It is not immediately clear why this theorem shows that chaos (or equivalently, superconcentration) implies multiple peaks. Let us prove that implication before starting the proof of Theorem 3.7. Recall the sequence $\mathbf{X}_n$ from Section 1 and the associated quantities. Suppose $v_n = o(\sigma_n^2)$. Let

$$
\alpha_n := \frac{v_n}{\sigma_n^2}.
$$

Put

$$
\epsilon_n := \alpha_n^{1/3}\sigma_n^2, \quad \delta_n := \alpha_n^{1/9}m_n, \quad l_n := [\alpha_n^{-1/24}].
$$

Then for sufficiently large n, a simple computation gives

$$
\begin{aligned}
\sqrt{\frac{v_n m_n l_n^3 \log l_n}{\delta_n \epsilon_n}} + \left(\frac{v_n \sigma_n^4 l_n^5 \log l_n}{\delta_n^3 m_n \epsilon_n}\right)^{1/4} &\leq \sqrt{\frac{v_n m_n l_n^4}{\delta_n \epsilon_n}} + \left(\frac{v_n \sigma_n^4 l_n^6}{\delta_n^3 m_n \epsilon_n}\right)^{1/4} \\
&\leq \alpha_n^{7/36} + \alpha_n^{1/48}\frac{\sigma_n}{m_n}.
\end{aligned}
$$

Clearly, $\epsilon_n = o(\sigma_n)$, $\delta_n = o(m_n)$, and $l_n \to \infty$. So the above bound implies multiple peaks as soon as we show that σ_n/m_n remains bounded. More is true, though. Let us now prove that $\sigma_n = o(m_n)$. Let x_+ denote the positive part of a real number x, and let

$$\psi(x) := \mathbb{E}(Z - x)_+^2,$$

where $Z \sim N(0,1)$. Clearly, ψ is everywhere positive and continuous, and $\lim_{x\to\infty} \psi(x) = 0$. For each n, suppose i_n is a coordinate such that $\mathrm{Var}(X_{n,i_n}) = \sigma_n^2$. Since surely $M_n \geq X_{n,i_n}$, we have

$$\begin{aligned} v_n = \mathbb{E}(M_n - m_n)^2 &\geq \mathbb{E}(M_n - m_n)_+^2 \\ &\geq \mathbb{E}(X_{n,i_n} - m_n)_+^2 = \sigma_n^2 \psi(m_n/\sigma_n). \end{aligned}$$

Since $v_n/\sigma_n^2 \to 0$, this shows that $m_n/\sigma_n \to \infty$, and completes our argument for chaos $\Rightarrow$ MP using Theorem 3.7.

Let us now give a sketch of the proof of Theorem 3.7. The idea is very simple: due to the chaos property, we can perturb the field a little bit to get a new maximum at a location 'far away' from the original maximum. However, since the perturbation is small and the value of the maximum is concentrated, it follows that the new location of the maximum must have been a near-maximal location in the unperturbed field. This shows the existence of at least two near-maximal points that are 'far away' from each other. Repeating the process as many times as we can, we get a large number of near-maxima.

Proof of Theorem 3.7. Let

$$t := \sqrt{\frac{v\delta \log l}{lm\epsilon}}.$$

Since $\delta < m$, we see that the quantity (9) is negative if $ltm/\delta > lt > 1$. Thus, we can assume without loss of generality that $lt \leq 1$. Then for any $1 \leq k \leq l$, we have the crude bound $1 - e^{-kt} \geq kt/2$. Thus, by Theorem 3.2,

$$(10) \qquad \sum_{k=1}^{l} \mathbb{E}(R(I^0, I^{kt})) \leq \sum_{k=1}^{l} \frac{v}{1 - e^{-kt}} \leq \sum_{k=1}^{l} \frac{2v}{kt} \leq \frac{2v \log l}{t}.$$

Let $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(l)}$ be i.i.d. copies of $\mathbf{X}$, and recursively define $\mathbf{Z}^{(0)}, \ldots, \mathbf{Z}^{(l)}$ as follows. Let $\mathbf{Z}^{(0)} := \mathbf{X}$, and for each k, let

$$\mathbf{Z}^{(k)} := e^{-t}\mathbf{Z}^{(k-1)} + \sqrt{1 - e^{-2t}}\mathbf{X}^{(k)}.$$

For each k, let

$$L_k := \mathrm{argmax}_{i \in S}\, Z_i^{(k)}.$$

It is easy to see by induction that $\mathbf{Z}^{(k)}$ has the same distribution as $\mathbf{X}$ for every k, and

$$\mathrm{Cov}(\mathbf{Z}^{(j)}, \mathbf{Z}^{(k)}) = e^{-t|j-k|}R.$$

This shows that

$$(\mathbf{Z}^{(j)}, \mathbf{Z}^{(k)}) \stackrel{dist.}{=} (\mathbf{X}^0, \mathbf{X}^{|j-k|t}).$$

In particular,

$$\mathbb{E}(R(L_j, L_k)) = \mathbb{E}(R(I^0, I^{|j-k|t})).$$

Thus, using (10) and the positivity of $R(i,j)$ we get that

$$\mathbb{P}\Big(\max_{1 \le j \ne k \le l} R(L_j, L_k) \ge \epsilon\Big) \le \frac{1}{\epsilon} \sum_{1 \le j \ne k \le l} \mathbb{E}(R(L_j, L_k))$$

(11)

$$\le \frac{l}{\epsilon} \sum_{k=1}^{l} \mathbb{E}(R(I^0, I^{kt})) \le \frac{2vl \log l}{\epsilon t}.$$

Now fix some $k \le l$. Note that $\mathbf{Z}^{(k)}$ can be written as

$$\mathbf{Z}^{(k)} = e^{-kt}\mathbf{X} + \sqrt{1 - e^{-2kt}}\mathbf{Y},$$

where $\mathbf{Y}$ is an independent copy of $\mathbf{X}$. Thus, if we let

$$\mathbf{W} := \sqrt{1 - e^{-2kt}}\mathbf{X} - e^{-kt}\mathbf{Y},$$

then $\mathbf{Z}^{(k)}$ and $\mathbf{W}$ are independent (because $\mathrm{Cov}(\mathbf{Z}^{(k)}, \mathbf{W}) = \mathbf{0}$), and

$$\mathbf{X} = e^{-kt}\mathbf{Z}^{(k)} + \sqrt{1 - e^{-2kt}}\mathbf{W}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}|X_{L_k} - m| &\le \mathbb{E}|e^{-kt}Z_{L_k}^{(k)} - m| + \sqrt{1 - e^{-2kt}}\mathbb{E}|W_{L_k}| \\
&\le e^{-kt}\mathbb{E}|Z_{L_k}^{(k)} - m| + (1 - e^{-kt})m + \sqrt{1 - e^{-2kt}}\mathbb{E}|W_{L_k}| \\
&\le \sqrt{v} + ktm + \sqrt{2kt}\sigma.
\end{aligned}$$

Note that we crucially used the independence of $\mathbf{Z}^{(k)}$ and $\mathbf{W}$ in the last line to conclude that $\mathbb{E}|W_{L_k}| \le \sigma$. From the above bound and Markov's inequality, we get

$$\begin{aligned}
\mathbb{P}\Big(\max_{1 \le k \le l} |X_{L_k} - m| \ge \tfrac{1}{2}\delta\Big) &\le \sum_{1 \le k \le l} \mathbb{P}\big(|X_{L_k} - m| \ge \tfrac{1}{2}\delta\big) \\
&\le \frac{2(l\sqrt{v} + l^2 tm + l^{3/2}\sigma\sqrt{2t})}{\delta}.
\end{aligned}$$

Again, applying Markov's inequality we have

$$\mathbb{P}(|M - m| \ge \tfrac{1}{2}\delta) \le \frac{2\sqrt{v}}{\delta}.$$

Combining the last two inequalities and the inequality (11), we get

$$\mathbb{P}\Big(\max_{1\leq j\neq k\leq l} R(L_j, L_k) \geq \epsilon \ \text{ or } \ \max_{1\leq k\leq l} |X_{L_k} - m| \geq \tfrac{1}{2}\delta \ \text{ or } \ |M - m| \geq \tfrac{1}{2}\delta\Big)$$

$$\leq \frac{2vl\log l}{\epsilon t} + \frac{2(l\sqrt{v} + l^2 tm + l^{3/2}\sigma\sqrt{2t})}{\delta} + \frac{2\sqrt{v}}{\delta}$$

$$= \sqrt{\frac{16vml^3\log l}{\delta\epsilon}} + \frac{(2l+2)\sqrt{v}}{\delta} + \left(\frac{4v\sigma^4 l^5\log l}{\delta^3 m\epsilon}\right)^{1/4}.$$

Now, since $\epsilon < \sigma^2$ and $l \geq 2$, the ratio of the second term to the third is bounded by

$$\left(\frac{64vm\epsilon}{\delta\sigma^4 l\log l}\right)^{1/4} \leq \left(\frac{64vm}{\delta\epsilon l\log l}\right)^{1/4} \leq \left(\frac{16vml^3\log l}{\delta\epsilon}\right)^{1/4},$$

which is precisely the square-root of the first term. Without loss of generality, we can assume that the first term is ≤ 1 (because we are bounding a probability), and therefore the third term dominates the second.

The proof is finished by defining $A := \{L_1, \ldots, L_l\}$. Note that although the construction of the set A involved auxiliary randomness, the *existence* of a set like A depends just on the vector $\mathbf{X}$, and hence the probability of existence is not affected when we expand the probability space. $\square$

3.4. Multiple Peaks does not imply Chaos.

In this subsection, we exhibit a sequence of Gaussian fields $(\mathbf{X}_n)_{n\geq 1}$ that satisfies the Multiple Peaks condition but is not chaotic. We will also arrange it so that the field has positive correlations between sites and the same variance at each site, just to show that these two factors don't play any role.

For simplicity, we fix n and avoid putting it as a subscript. Let

$$(g_{ij}^k, 1 \leq i \leq n, \ j \in \{0,1\}, \ 1 \leq k \leq n)$$

be a collection of i.i.d. standard Gaussian random variables. Let T be the set of all maps from $\{1, \ldots, n\}$ into $\{0,1\}$. For each $f \in T$ and each $k \leq n$, define

$$Y_f^k := \frac{\sum_{i=1}^n g_{if(i)}^k}{\sqrt{n}}.$$

Let $(Z_k)_{k\leq n}$ be i.i.d. standard Gaussian random variables, and let

$$\rho := 1 - n^{-1/3}.$$

Finally, define the vector $\mathbf{X} = (X_f^k)_{f\in T, \, k\leq n}$ as

$$X_f^k = \begin{cases} Y_f^k & \text{if } k = 1, \\ \rho Y_f^k + \sqrt{1-\rho^2} Z_k & \text{if } k > 1. \end{cases}$$

Then note that for each $f, f' \in T$ and $k, k' \leq n$,

$$\mathrm{Var}(X_f^k) = 1, \quad R((k,f),(k',f')) := \mathrm{Cov}(X_f^k, X_{f'}^{k'}) \geq 0.$$

Thus, in our notation, we have

$$\sigma_n^2 = \max_{f \in T,\ k \le n} \mathrm{Var}(X_f^k) = 1.$$

We claim that $\mathbf{X}$ has multiple peaks but is not chaotic. First, we show the existence of multiple peaks. Observe that for any $k \le n$,

$$\tag{12} \max_f Y_f^k = \frac{\sum_{i=1}^n \max\{g_{i0}, g_{i1}\}}{\sqrt{n}}.$$

Thus, if

$$\mu := \mathbb{E} \max\{g_{10}, g_{11}\},$$

and

$$M_k := \max_f X_f^k,$$

then we have

$$\mathbb{E}(M_k) = \begin{cases} \mu\sqrt{n} & \text{if } k = 1, \\ \rho\mu\sqrt{n} & \text{if } k > 1. \end{cases}$$

Note that

$$\max_k M_k = \max_{f \in T,\ 1 \le k \le n} X_f^k =: M.$$

In particular,

$$m_n := \mathbb{E}(M) \ge \mu\sqrt{n}.$$

By Proposition 1.3, for each k we have

$$\mathbb{P}(|M_k - \mathbb{E}(M_k)| \ge 2\mu n^{1/6}) \le 2e^{-2\mu^2 n^{1/3}}.$$

Thus,

$$\mathbb{P}(\max_k |M_k - \mathbb{E}(M_k)| \ge 2\mu n^{1/6}) \le 2ne^{-2\mu^2 n^{1/3}}.$$

Now,

$$\max_k |M_k - \mu\sqrt{n}| \le \max_k |M_k - \mathbb{E}(M_k)| + \max_k |\mathbb{E}(M_k) - \mu\sqrt{n}|$$
$$\le \max_k |M_k - \mathbb{E}(M_k)| + \mu n^{1/6}.$$

Therefore,

$$\tag{13} \mathbb{P}\big(\max_k |M_k - \mu\sqrt{n}| \ge 3\mu n^{1/6}\big) \le 2ne^{-2\mu^2 n^{1/3}}.$$

Now choose

$$A = \{(1, f_1), (2, f_2), \dots, (n, f_n)\},$$

where f_k is the f that maximizes X_f^k. By (13), if we take $\gamma_n = 2ne^{-2\mu^2 n^{1/3}}$, then with probability $\ge 1 - \gamma_n$, we have

$$\text{for all } k, \quad X_{f_k}^k = M_k \ge M - 6\mu n^{1/6}.$$

Also note that $|A| = n$ and $R((i, f_i), (j, f_j)) = 0$ for all $i \ne j$. Thus, the multiple peaks condition holds with $l_n = n$, $\epsilon_n = 0$, $\delta_n = 6\mu n^{1/6}$, and γ_n as above. Clearly, $l_n \to \infty$, $\epsilon_n = o(\sigma_n^2)$, $\delta_n = o(m_n)$, and $\gamma_n \to 0$.

Next, let us show that $\mathbf{X}$ is not chaotic. We accomplish this by showing the $\mathbf{X}$ is not superconcentrated. Since

$$\rho\mu\sqrt{n} = \mu\sqrt{n} - \mu n^{1/6},$$

Proposition 1.3 implies that

$$\mathbb{P}(M_1 < M) \leq \mathbb{P}\left(M_1 < \mu\sqrt{n} - \frac{\mu}{2}n^{1/6}\right)$$
$$+ \mathbb{P}\left(\max_{2 \leq k \leq n} M_k > \rho\mu\sqrt{n} + \frac{\mu}{2}n^{1/6}\right)$$
$$\leq ne^{-\frac{1}{8}\mu^2 n^{1/3}}.$$

Therefore,

$$v_n := \mathrm{Var}(M) \geq \mathbb{E}((M - \mathbb{E}(M))^2; M_1 = M)$$
$$= \mathbb{E}(M_1 - \mathbb{E}(M))^2 - \mathbb{E}((M_1 - \mathbb{E}(M))^2; M_1 < M)$$
$$\geq \mathrm{Var}(M_1) - \left[\mathbb{E}(M_1 - \mathbb{E}(M))^4 \mathbb{P}(M_1 < M)\right]^{1/2}.$$

From (12) we see that $\mathrm{Var}(M_1) = \mathrm{Var}(\max\{g_{10}, g_{11}\})$. From Proposition 1.3, it follows that $\mathbb{E}(M_1 - \mathbb{E}(M_1))^4 \leq C$, where C is a universal constant. Again, by an argument similar to Lemma 2.1 (using the Gaussian tail bounds for $M_1, \ldots, M_n$ that we get by Proposition 1.3), we can show that $|\mathbb{E}(M_1) - \mathbb{E}(M)| \leq C\sqrt{\log n}$. Combining these observations and the exponentially decaying bound on $\mathbb{P}(M_1 < M)$ obtained above, we see that v_n can be bounded below by a positive constant that does not depend on n. Since $\sigma_n^2 \equiv 1$, this completes the argument.

3.5. Chaos does not imply Strong Multiple Peaks.
In this section, we produce a sequence of Gaussian fields that is superconcentrated, but does not have multiple peaks in the strong sense. As in the previous section, we fix n and avoid putting it as a subscript. Let $(g_{ij})_{1 \leq i,j \leq n}$ be a collection of i.i.d. standard Gaussian random variables. Let $\mathcal{F}$ be the set of all functions from $\{1, \ldots, n\}$ into itself. For each $f \in \mathcal{F}$, let

$$X_f := \frac{\sum_{i=1}^n g_{if(i)}}{\sqrt{n}}.$$

Then $\mathbf{X} := (X_f)_{f \in \mathcal{F}}$ is a centered Gaussian random vector, with $\mathbb{E}(X_f^2) = 1$ for each f, and for any f, f',

$$R(f, f') := \mathrm{Cov}(X_f, X_{f'}) = \frac{|\{i : f(i) = f'(i)\}|}{n} \in [0, 1].$$

Now, if we let

$$M_i := \max_{1 \leq j \leq n} g_{ij},$$

then clearly,

$$M := \max_f X_f = \frac{\sum_{i=1}^n M_i}{\sqrt{n}}.$$

Since $M_1, \ldots, M_n$ are i.i.d. and $\mathrm{Var}(M_i) \leq C/\log n$ for some universal constant C (well-known result; see Proposition 4.2), therefore we have

$$\mathrm{Var}(M) \leq \frac{C}{\log n}.$$

This shows that $\mathbf{X}$ is superconcentrated. Let us now show that multiple peaks are not present in the strong sense. We need the following simple lemma about the difference between the largest and second largest values in a realization of n i.i.d. standard Gaussian random variables.

Lemma 3.8. *Suppose $Z_1, \ldots, Z_n$ are i.i.d. standard Gaussian random variables. Let D denote the difference between the largest and second largest of these values. There is a constant $\rho > 0$ such that for any n,*

$$\mathbb{P}\left(D > \frac{\rho}{\sqrt{\log n}}\right) > \frac{7}{8}.$$

Proof. Let M_n and M_n^- denote the largest and second largest values among $Z_1, \ldots, Z_n$. Let

$$b_n = \sqrt{2 \log n} - \frac{\log \log n + \log(4\pi)}{2\sqrt{2 \log n}}, \quad a_n = \sqrt{2 \log n}.$$

Let $\Phi(x) = \mathbb{P}(Z_1 \leq x)$. Then clearly, for any $-\infty < y < x < \infty$,

$$\mathbb{P}(M_n^- \leq x, \ M_n > y) = n(1 - \Phi(y))\Phi(x)^{n-1}.$$

Using this and the Mills ratio bounds from Section 2, it is easy to show that for any $-\infty < u < v < \infty$,

$$\lim_{n \to \infty} \mathbb{P}\big(a_n(M_n^- - b_n) \leq u, \ a_n(M_n - b_n) > v\big) = e^{-u}e^{-e^{-v}}.$$

From this, it can deduced that $a_n(M_n - M_n^-)$ converges in law to a distribution that has a positive density on $(0, \infty)$ and no point masses. This proves the lemma. $\qquad\square$

Let us now finish the proof of the nonexistence of strong multiple peaks in the field $\mathbf{X}$. For each i, let M_i^- be the second largest value among $g_{i1}, \ldots, g_{in}$. Let $D_i := M_i - M_i^-$. Let $D_{(1)} \leq D_{(2)} \leq \cdots \leq D_{(n)}$ denote the D_i's arranged in increasing order. Let ρ be the constant from the above lemma. Then by the lemma and the weak law of large numbers we see that as $n \to \infty$,

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{1}{n}\left|\left\{i : D_i > \frac{\rho}{\sqrt{\log n}}\right\}\right| \geq \frac{7}{8}\right) = 1.$$

(Note that we should have written $D_{n,i}$ instead of D_i, but we will be slack about this kind of thing.) It is not difficult to see from here that

$$(14) \qquad \lim_{n \to \infty} \mathbb{P}\left(\sum_{i=1}^{[n/4]} D_{(i)} \geq \frac{\rho n}{8\sqrt{\log n}}\right) = 1.$$

Let f^* denote the maximizing function, that is $f^*(i) = \mathrm{argmax}_j g_{ij}$. Suppose f, f' are two functions such that $R(f, f') < 1/2$. Then we must have that

$\min\{R(f, f^*), R(f', f^*)\} < 3/4$, because if $R(f, f^*) \geq 3/4$ and $R(f', f^*) \geq 3/4$, then $R(f, f') \geq 1/2$. Thus,

$$(15) \qquad X_{f^*} - \min\{X_f, X_{f'}\} \geq \frac{\sum_{i=1}^{[n/4]} D_{(i)}}{\sqrt{n}}.$$

Let E denote the following event: For any $f, f' \in \mathcal{F}$ such that $R(f, f') < 1/2$, at least one of X_f and $X_{f'}$ is less than

$$X_{f^*} - \frac{\rho}{8}\sqrt{\frac{n}{\log n}}.$$

Then from (14) and (15), we see that $\mathbb{P}(E) \to 1$ as $n \to \infty$. This shows that the Strong Multiple Peaks condition does not hold for the sequence of Gaussian fields under consideration.

4. Hypercontractivity

Suppose we have a semigroup of operators $(P_t)_{t \geq 0}$, acting on some space of functions on $\mathbb{R}^n$ (semigroup means $P_t P_s = P_{t+s}$). Suppose μ is an invariant measure for the semigroup, meaning that $\int P_t f \, d\mu = \int f \, d\mu$ for any f in the domain of P_t. The semigroup is said to be 'hypercontractive' if for any $t > 0$, there are numbers $q > p > 1$ (possibly depending on t) such that P_t maps $L^q(\mu)$ into $L^p(\mu)$ and moreover, $\|P_t f\|_q \leq \|f\|_p$ for all $f \in L^q(\mu)$. Here $\|\cdot\|_p$ denotes the standard L^p-norm on $L^p(\mu)$. This phenomenon was discovered by Nelson [56] and established for the Ornstein-Uhlenbeck semigroup by him a few years later [57]. The O-U semigroup on $\mathbb{R}^n$ is defined as follows. Let $\mathbf{Z}$ be a standard Gaussian random vector in $\mathbb{R}^n$. For any function f and $t \geq 0$, define the function $P_t f$ as

$$P_t f(\mathbf{x}) := \mathbb{E} f(e^{-t}\mathbf{x} + \sqrt{1 - e^{-2t}}\mathbf{Z}).$$

It is easy to see that the standard Gaussian measure on $\mathbb{R}^n$, which we denote by $\gamma^{\otimes n}$, is an invariant measure for this semigroup. Nelson [57] showed that for any $p > 1$ and $t \geq 0$, if we let $q = 1 + e^{2t}(p - 1) > p$, then for all $f \in L^q(\gamma^{\otimes n})$, we have

$$(16) \qquad \|P_t f\|_q \leq \|f\|_p.$$

Note that the result does not depend on the dimension n at all. This is one of the remarkable properties that make hypercontractivity a deep and powerful tool.

The study of hypercontractive semigroups was given a major boost by the discovery of the connection between hypercontractivity and logarithmic Sobolev inequalities by Gross [31]. For surveys of the extensive literature that developed around this topic, one can look in the wonderful monographs [4] and [35].

The connection between logarithmic Sobolev inequalities (and hence hypercontractivity) and concentration of measure was discovered by I. Herbst in a small unpublished note (see Ledoux [47], Theorem 5.3). However, the

Herbst argument can only establish ordinary concentration of measure. The first application of hypercontractivity to prove what we call superconcentration was due to Talagrand [71]. Talagrand's method was subsequently used by Benjamini, Kalai, and Schramm [7] to prove superconcentration in first passage percolation (they call it 'sublinear variance').

The way we use hypercontractivity in this paper to establish superconcentration is essentially the same in spirit as Talagrand's technique, although there is the difference that while Talagrand's method applies only to functions of independent random variables, we can deal with strong dependence as in the Gaussian free field.

Recall that in our setting, we have a vector $\mathbf{X} = (X_i)_{i \in S}$ that is centered Gaussian but not necessarily with independent components. We can still define a semigroup $(\tilde{P}_t)_{t \geq 0}$ as

$$\tilde{P}_t f(\mathbf{x}) := \mathbb{E} f(e^{-t}\mathbf{x} + \sqrt{1 - e^{-2t}}\mathbf{X}).$$

Note that this semigroup retains the same hypercontractive property (16) as the standard O-U semigroup, with exactly the same relation between p and q. This can be argued as follows. In our notation, $R = (R(i,j))_{i,j \in S}$ is the covariance matrix of $\mathbf{X}$. Assuming that the vector $\mathbf{X}$ is not identically equal to zero, we know that for some $d \leq |S|$, there is a $|S| \times d$ matrix B of full rank such that $BB^T = R$. We can assume that there is a d-dimensional standard Gaussian vector $\mathbf{Z}$ such that $\mathbf{X} = B\mathbf{Z}$. Let $\mathbf{Z}'$ be an independent copy of $\mathbf{Z}$, and let $\mathbf{X}' = B\mathbf{Z}'$. Given $f : \mathbb{R}^S \to \mathbb{R}$, define the function $g : \mathbb{R}^d \to \mathbb{R}$ as

$$g(\mathbf{x}) := f(B\mathbf{x}).$$

The random vector $\mathbf{X}$ is supported on the image of $\mathbb{R}^d$ under the map B. On this subspace the map B has an inverse, which we denote by B^{-1}. Then we have

$$\begin{aligned}
\tilde{P}_t f(\mathbf{x}) &= \mathbb{E} f(e^{-t}\mathbf{x} + \sqrt{1 - e^{-2t}}\mathbf{X}) \\
&= \mathbb{E} f(e^{-t}BB^{-1}\mathbf{x} + \sqrt{1 - e^{-2t}}B\mathbf{Z}) \\
&= \mathbb{E} g(e^{-t}B^{-1}\mathbf{x} + \sqrt{1 - e^{-2t}}\mathbf{Z}) = P_t g(B^{-1}\mathbf{x}).
\end{aligned}$$

Therefore, we have that for every $t \geq 0$,

$$\tilde{P}_t f(\mathbf{X}) = P_t g(B^{-1}\mathbf{X}) = P_t g(\mathbf{Z}),$$

and so by (16),

$$\|\tilde{P}_t f(\mathbf{X})\|_q = \|P_t g(\mathbf{Z})\|_q \leq \|g(\mathbf{Z})\|_p = \|g(\mathbf{X})\|_p.$$

Thus, the hypercontractive property (16) holds for the semigroup $(\tilde{P}_t)_{t \geq 0}$ with the same q, p. The following lemma, which is of crucial importance to us, is a direct consequence. Recall all notation from Section 1, including the definition of $\mathbf{X}^t$.

Lemma 4.1. *For any measurable $f : \mathbb{R}^S \to [0,1]$ and any $t \geq 0$, we have*

$$\mathbb{E}(f(\mathbf{X})f(\mathbf{X}^t)) \leq (\mathbb{E}f(\mathbf{X}))^{1+\tanh(t/2)}.$$

Proof. Take any $p > 1$ and let $q = 1 + e^{2t}(p - 1)$. Let $q' = q/(q - 1)$. Since f maps into $[0, 1]$, we have

$$
\begin{aligned}
\mathbb{E}(f(\mathbf{X})f(\mathbf{X}^t)) = \mathbb{E}(f(\mathbf{X})\tilde{P}_t f(\mathbf{X})) &\leq \|f(\mathbf{X})\|_{q'}\|\tilde{P}_t f(\mathbf{X})\|_q \\
&\leq \|f(\mathbf{X})\|_{q'}\|f(\mathbf{X})\|_p \\
&\leq \left(\mathbb{E}f(\mathbf{X})\right)^{\frac{1}{q'}+\frac{1}{p}}.
\end{aligned}
$$

We now optimize over p, which yields

$$
p = 1 + e^{-t}.
$$

An easy computation gives

$$
\begin{aligned}
\frac{1}{q'} + \frac{1}{p} &= 1 - \frac{1}{1 + e^{2t}(p - 1)} + \frac{1}{p} \\
&= 1 + \tanh(t/2).
\end{aligned}
$$

This completes the proof. $\qquad\square$

Lemma 4.1 is a very useful tool, as we will see in our proof of superconcentration in directed polymers in Section 8 and in the Kauffman-Levin NK model in Section 7. It is particularly potent in combination with Lemma 3.3 or Theorem 3.1. To demonstrate an immediate application, let us work out the following simple result that we do not know how to prove without using hypercontractivity. It says that the variance of the maximum is small whenever the correlations are uniformly small. Recall that $v = \text{Var}(\max_i X_i)$ and $R(i, j) = \text{Cov}(X_i, X_j)$.

Proposition 4.2. *Suppose that $R(i, j) \leq \rho$ for each $i \neq j$ and $R(i, i) = \sigma^2$ for each i. Then*

$$
v \leq \frac{C\sigma^2}{\log |S|} + C\rho,
$$

where C is a universal constant.

Remark. It can be easily shown that the bound is sharp by considering the case where $R(i, j) = \rho$ for each $i \neq j$. However, in spite of the sharpness, the result is probably not very useful since the uniform boundedness is a very strong restriction. It is presented here only for illustration purposes.

Proof. For any vector $\mathbf{x} \in \mathbb{R}^S$ all of whose coordinates have distinct values, define

$$
f_i(\mathbf{x}) := \mathbb{I}_{\{x_i = \max_j x_j\}}.
$$

Then by Lemma 4.1 and the hypothesis of this theorem, we have

$$
\begin{aligned}
\mathbb{E}(R(I^0, I^t)) &\leq \sigma^2 \mathbb{P}(I^0 = I^t) + \rho \mathbb{P}(I^0 \neq I^t) \\
&= \sigma^2 \sum_{i \in S} \mathbb{E}(f_i(\mathbf{X}^0) f_i(\mathbf{X}^t)) + \rho \mathbb{P}(I^0 \neq I^t) \\
&\leq \sigma^2 \sum_{i \in S} (\mathbb{E} f_i(\mathbf{X}^0))^{1+\tanh(t/2)} + \rho \\
&\leq \sigma^2 (\max_{i \in S} \mathbb{P}(I^0 = i))^{\tanh(t/2)} \sum_{i \in S} \mathbb{P}(I^0 = i) + \rho \\
&= \sigma^2 (\max_{i \in S} \mathbb{P}(I^0 = i))^{\tanh(t/2)} + \rho.
\end{aligned}
$$

Now, by increasing the value of the constant in the statement of the theorem (and recalling Proposition 1.2), we can assume without loss of generality that $\rho \leq \sigma^2/2$. Under this assumption, by the Sudakov minoration technique stated in Section 2, we have

$$
m \geq C\sigma \sqrt{\log |S|},
$$

where we are using our convention of letting C denote any universal constant. Thus, by Proposition 1.3 and the tail bound (2) we have

$$
\begin{aligned}
\mathbb{P}(I^0 = i) &\leq \mathbb{P}(X_i \geq m/2) + \mathbb{P}(M \leq m/2) \\
&\leq 2e^{-m^2/8\sigma^2} \leq |S|^{-C}.
\end{aligned}
$$

Now, for each $t \geq 0$,

$$
(17) \qquad \tanh(t/2) = \frac{e^{t/2} - e^{-t/2}}{e^{t/2} + e^{-t/2}} = \frac{(1 - e^{-t/2})(1 + e^{-t/2})}{1 + e^{-t}} \geq 1 - e^{-t/2}.
$$

Combining this with the bounds on $\mathbb{E}(R(I^0, I^t))$ and $\mathbb{P}(I^0 = i)$ obtained above and Theorem 3.1, we get

$$
\begin{aligned}
v &= \int_0^\infty e^{-t} \mathbb{E}(R(I^0, I^t)) dt \\
&\leq \int_0^\infty e^{-t} \big(\sigma^2 |S|^{-C(1-e^{-t/2})} + \rho \big) dt \\
&\leq \int_0^\infty e^{-t/2} \sigma^2 |S|^{-C(1-e^{-t/2})} dt + \rho \\
&= 2\sigma^2 \int_0^1 |S|^{-Cu} du + \rho \leq \frac{C\sigma^2}{\log |S|} + \rho.
\end{aligned}
$$

This completes the proof.   □

A part of the proof of Proposition 4.2 generalizes to the following technique, that will be help us bound the variance of the maximum of a discrete Gaussian free field in Section 10. It will also be used for analyzing continuous Gaussian fields on Euclidean spaces in Section 11.

Theorem 4.3. *Suppose that for each $r \geq 0$, there is a covering $\mathcal{C}(r)$ of S such that whenever i, j are indices with $R(i, j) \geq r$, we have $i, j \in D$ for some $D \in \mathcal{C}(r)$. For each $A \subseteq S$ let $p(A) := \mathbb{P}(I^0 \in A)$, and define*

$$\rho(r) := \max_{D \in \mathcal{C}(r)} p(D), \quad \mu(r) := \sum_{D \in \mathcal{C}(r)} p(D) = \mathbb{E}\big|\{D \in \mathcal{C}(r) : I^0 \in D\}\big|.$$

Then we have

$$v \leq \int_0^{\sigma^2} \frac{2\mu(r)(1 - \rho(r))}{-\log \rho(r)} \, dr,$$

where we interpret $(1 - x)/\log x = -1$ when $x = 1$.

Proof. By Lemma 4.1, we have

$$\mathbb{P}(R(I^0, I^t) \geq r) \leq \sum_{D \in \mathcal{C}(r)} \mathbb{P}(I^0 \in D, \ I^t \in D)$$

$$\leq \sum_{D \in \mathcal{C}(r)} p(D)^{1+\tanh(t/2)}$$

$$\leq \rho(r)^{\tanh(t/2)} \sum_{D \in \mathcal{C}(r)} p(D) = \mu(r)\rho(r)^{\tanh(t/2)}.$$

Thus, by Lemma 3.1 we have

$$v = \int_0^\infty e^{-t} \mathbb{E}(R(I^0, I^t)) \, dt$$

$$\leq \int_0^\infty \int_0^{\sigma^2} e^{-t} \mathbb{P}(R(I^0, I^t) \geq r) \, dr \, dt$$

$$\leq \int_0^{\sigma^2} \int_0^\infty e^{-t} \mu(r)\rho(r)^{\tanh(t/2)} \, dt \, dr.$$

Now, by (17) we have that for any fixed r,

$$\int_0^\infty e^{-t} \rho(r)^{\tanh(t/2)} \, dt \leq \int_0^\infty e^{-t} \rho(r)^{1-e^{-t/2}} \, dt$$

$$\leq \int_0^\infty e^{-t/2} \rho(r)^{1-e^{-t/2}} \, dt$$

$$= \int_0^1 2\rho(r)^u \, du = \frac{2(1 - \rho(r))}{-\log \rho(r)}.$$

This completes the proof. $\qquad\square$

5. EXTREMAL FIELDS

The goal of this section is to prove Theorems 1.10 and 1.11. As usual, we prove quantitative versions. Recall the definitions of $\mathbf{X}, \mathbf{X}^t, M, m, v, \sigma^2, I^t$, and $R(i, j)$ from Section 1. The following theorem is the main result of this section.

Theorem 5.1. *Let*

$$\alpha := \frac{m}{\sqrt{2\sigma^2 \log |S|}}, \quad \beta := \sqrt{1-\alpha} + \left(\frac{\log\log |S|}{\log |S|}\right)^{1/4}.$$

Then $v \le C\sigma^2\beta$ and for any $t \ge 0$, $\mathbb{E}(R(I^0, I^t)) \le \frac{C\sigma^2\beta}{\sqrt{1-e^{-2t}}}$, where C is a universal constant.

It is clear how Theorem 1.10 follows from this result. However, Theorem 1.11 still needs a proof. Let us prove it before moving on to prove Theorem 5.1. As usual C will denote any universal constant, whose value may change from line to line.

Proof of Theorem 1.11. Without loss of generality, assume that $\sigma^2 = 1$. Let $n = |S|$, and define

$$N := |\{i : X_i \ge \sqrt{2\log n}\}|.$$

Then by the Mills ratio lower bound from Section 2 and the assumption that $R(i,i) = 1$ for all i, we have

$$\mathbb{E}(N) \ge Cn\frac{e^{-\log n}}{\sqrt{2\log n}} = \frac{C}{\sqrt{2\log n}}.$$

Again, we have

$$\begin{aligned}
\mathbb{E}(N^2) &= \sum_{i,j\in S} \mathbb{P}\big(X_i \ge \sqrt{2\log n}, \; X_j \ge \sqrt{2\log n}\big) \\
&\le \sum_{i,j\in S} \mathbb{P}\big(X_i + X_j \ge 2\sqrt{2\log n}\big) \\
&\le \sum_{i,j\in S} \exp\left(-\frac{4\log n}{\mathrm{Var}(X_i + X_j)}\right) \\
&\le \sum_{i,j\in S} \exp\left(-\frac{2\log n}{1 + R(i,j)}\right).
\end{aligned}$$

Applying the Paley-Zygmund second moment method, we get

$$\mathbb{P}\big(M \ge \sqrt{2\log n}\big) = \mathbb{P}(N > 0) \ge \frac{(\mathbb{E}(N))^2}{\mathbb{E}(N^2)} \ge \frac{C}{\sum_{i,j\in S} n^{-2/(1+R(i,j))}\log n}.$$

Again, by Proposition 1.3 we know that for any $x \ge 0$,

$$\mathbb{P}(M - \mathbb{E}(M) \ge x) \le e^{-x^2/2}.$$

Combining, we see that

$$\mathbb{E}(M) \ge \sqrt{2\log n} - C\left(\log\log n + \log \sum_{i,j\in S} n^{-2/(1+R(i,j))}\right)^{1/2}.$$

The proof now follows from Theorem 5.1. $\qquad\square$

Let us now give a brief idea of the proof of Theorem 5.1 before going into the technical details. The purpose of this sketch is to succinctly convey an idea that may have the potential to grow as an alternative to the hypercontractive method for proving superconcentration. We begin by assuming that $\sigma^2 = 1$. The first step is to show that for any $t \geq 0$, $X_{I^t} \approx e^{-t}m$ with high probability. We carry out this important step in Lemma 5.2. Next, we fix t and let $D = \{i : X_i \approx e^{-t}m\}$, where the meaning of $\approx$ has to be made precise. A simple first moment bound shows that with high probability, $|D| \lesssim n^{1-e^{-2t}}$. Next we fix some $r \geq 0$ and let $B = \{i : R(I_0, i) \geq r\}$. The key observation is that since $\mathbf{X}'$ and $\mathbf{X}$ are independent and $\sigma^2 = 1$, we have $\mathrm{Var}(X_i' - rX_{I^0}' \mid \mathbf{X}) \leq 1 - r^2$ for each $i \in B$, and hence

$$\mathbb{E}\big(\max_{i \in B \cap D} X_i' \mid \mathbf{X}\big) = \mathbb{E}\big(\max_{i \in B \cap D}(X_i' - rX_{I^0}') \mid \mathbf{X}\big) \lesssim \sqrt{(1-r^2)2\log|D|}.$$

Combining this with the bound $|D| \lesssim n^{1-e^{-2t}}$, we get

$$\max_{i \in B \cap D} X_i^t \leq e^{-t} \max_{i \in B \cap D} X_i + \sqrt{1 - e^{-2t}} \max_{i \in B \cap D} X_i'$$
$$\lesssim e^{-2t}\alpha\sqrt{2\log n} + (1 - e^{-2t})\sqrt{2(1-r^2)\log n}.$$

Thus, if $\sqrt{1-r^2} < \alpha$, or in other words $r > \sqrt{1-\alpha^2}$, we cannot have that $\max_{i \in B \cap D} X_i^t \approx \alpha\sqrt{2\log n}$. If this does not happen, then $I^t \notin B \cap D$. But we have already stated that $I^t \in D$ with high probability. Therefore, whenever $r > \sqrt{1-\alpha^2}$, we have that $I^t \in D \backslash B$ with high probability, which implies that $R(I^0, I^t) < r$ with high probability. Roughly, this justifies the $\sqrt{1-\alpha}$ term in the statement of the theorem. The second term arises from our attempts at making the above sketch rigorous, which is a somewhat technically involved task.

Let us now begin the formal proof. The first step, as mentioned before, is to show that $X_{I^t} \approx e^{-t}m$. This is made precise in the following lemma.

Lemma 5.2. *Take any $t \geq 0$. Then for any $x \geq 0$, we have*

$$\mathbb{P}\big(|X_{I^t} - e^{-t}m| \geq x\big) \leq 4e^{-x^2/4\sigma^2}.$$

Proof. For notational convenience, let $a = e^{-t}$, $b = \sqrt{1-e^{-2t}}$, $\mathbf{Z} := \mathbf{X}^t$, and $\mathbf{W} := b\mathbf{X} - a\mathbf{X}'$. Then $\mathbf{Z}$ and $\mathbf{W}$ are independent, and

(18) $$\mathbf{X} = a\mathbf{Z} + b\mathbf{W}.$$

Since $a + b \leq \sqrt{2}$, by Proposition 1.3 and the independence of $\mathbf{Z}$ and $\mathbf{W}$, we have

$$\mathbb{P}(|X_{I^t} - am| \geq x) \leq \mathbb{P}(a|Z_{I^t} - m| \geq ax/\sqrt{2}) + \mathbb{P}(b|W_{I^t}| \geq bx/\sqrt{2})$$
$$\leq 2e^{-x^2/4\sigma^2} + 2e^{-x^2/4\sigma^2}.$$

This completes the proof. $\square$

Proof of Theorem 5.1. Without loss of generality, we assume that $\sigma^2 = 1$. As in Lemma 5.2, let $a = e^{-t}$, $b = \sqrt{1-e^{-2t}}$, and $\mathbf{Z} = \mathbf{X}^t$. Let $n = |S|$.

Note that since $|R(i,j)| \leq 1$ for all i,j and C can be chosen as large as we like, it suffices to prove the theorem assuming that n is larger than some fixed threshold and that

$$(19) \qquad \frac{1}{b}\left(\frac{\log\log n}{\alpha^2 \log n}\right)^{1/4} \leq \frac{1}{100}.$$

For the same reason, there is no loss of generality in assuming that $m \geq 2$. With that assumption, define

$$(20) \qquad \delta := \frac{\sqrt{\log(m^4/2)}}{m} \in (0,1),$$

and let

$$D := \{i : |X_i - am| \leq \delta m\}.$$

By Lemma 5.2,

$$(21) \qquad \mathbb{P}(I^t \notin D) \leq 4e^{-\delta^2 m^2/4}.$$

Now, if $a \geq \delta$, then by the Mills ratio upper bound from Section 2,

$$\mathbb{E}|D| \leq \sum_i \mathbb{P}(X_i \geq (a-\delta)m)$$

$$\leq ne^{-(a-\delta)^2 m^2/2} = n^{1-(a-\delta)^2\alpha^2}.$$

Therefore, if we define

$$\zeta := \begin{cases} 1 - (a-\delta)^2\alpha^2 + \delta^2 & \text{if } a \geq \delta, \\ 1 & \text{if } a < \delta, \end{cases}$$

then in all cases we have

$$(22) \qquad \mathbb{P}(|D| > n^\zeta) \leq n^{-\delta^2}.$$

Define

$$\gamma := \frac{\alpha(b^2 - 2\delta)}{b\sqrt{\zeta}}.$$

It is easy to verify using (19) that $b^2 \geq 2\delta$ and hence $\gamma \geq 0$. Again,

$$\frac{\alpha(b^2-2\delta)}{b\sqrt{\zeta}} \leq \frac{b^2-2\delta}{b\sqrt{1-a^2}} = \frac{b^2-2\delta}{b^2} \leq 1.$$

Thus, $\gamma \in [0,1]$. Let $r := \sqrt{1-\gamma^2}$, and define the random set

$$B := \{i : R(I^0, i) \geq r\}.$$

Note that

$$(23) \qquad \mathbb{P}(R(I^0, I^t) \geq r) = \mathbb{P}(I^t \in B) \leq \mathbb{P}(I^t \in B \cap D) + \mathbb{P}(I^t \notin D).$$

Let $\mathbb{E}^0$ and Var^0 denote the conditional expectation and variance given $\mathbf{X}$. Since $R(i,i) \leq 1$ and $R(I^0, i) \geq r$ for all $i \in B$ and $\mathbf{X}'$ is independent of $\mathbf{X}$, we have

$$\text{Var}^0(X'_i - rX'_{I^0}) \leq 1 + r^2 - 2r^2 = \gamma^2.$$

Again, $\mathbb{E}^0(X'_{I^0}) = 0$. Thus, if $B \cap D \neq \emptyset$, then by Lemma 2.1 we have

$$\mathbb{E}^0(\max_{i \in B \cap D} X'_i) = \mathbb{E}^0(\max_{i \in B \cap D}(X'_i - rX'_{I^0}))$$

$$\leq \gamma\sqrt{2\log|B \cap D|} \leq \gamma\sqrt{2\log|D|}.$$

Combined with Proposition 1.3, this implies that if $B \cap D \neq \emptyset$, then for all $x \geq 0$,

$$(24) \qquad \mathbb{P}^0(\max_{i \in B \cap D} X'_i \geq \gamma\sqrt{2\log|D|} + x) \leq e^{-x^2/2}.$$

Clearly, the inequality holds even if we relax the condition $B \cap D \neq \emptyset$ to just $D \neq \emptyset$, interpreting the maximum of an empty set as $-\infty$. Since $X_i \leq (a + \delta)m$ for $i \in D$ and $\mathbf{Z} = a\mathbf{X} + b\mathbf{X}'$, we have

$$\max_{i \in B \cap D} Z_i \leq a\max_{i \in B \cap D} X_i + b\max_{i \in B \cap D} X'_i$$

$$\leq a(a + \delta)m + b\max_{i \in B \cap D} X'_i.$$

Thus, from (24), we see that whenever $D \neq \emptyset$,

$$(25) \qquad \mathbb{P}^0(\max_{i \in B \cap D} Z_i \geq a(a + \delta)m + b\gamma\sqrt{2\log|D|} + bx) \leq e^{-x^2/2}.$$

Putting

$$m' := a(a + \delta)m + b\gamma\sqrt{2\zeta\log n}$$

$$= \left(a(a + \delta) + \frac{b\gamma\sqrt{\zeta}}{\alpha}\right)m$$

$$= (a^2 + a\delta + b^2 - 2\delta)m = (1 + a\delta - 2\delta)m,$$

and using (25) and (22) we get

$$\mathbb{P}(\max_{i \in B \cap D} Z_i \geq m' + bx) \leq \mathbb{E}(\mathbb{P}^0(\max_{i \in B \cap D} Z_i \geq m' + bx); 1 \leq |D| \leq n^\zeta)$$

$$(26) \qquad\qquad\qquad + \mathbb{E}(\mathbb{P}^0(\max_{i \in B \cap D} Z_i \geq m' + bx); |D| > n^\zeta)$$

$$\leq e^{-x^2/2} + n^{-\delta^2}.$$

Now

$$(27) \qquad\qquad m - m' = (2 - a)\delta m \geq \delta m.$$

In particular, $m' \leq m$. Let $x := (m - m')/2b$. Then by (26) and Proposition 1.3 we have

$$\mathbb{P}(I^t \in B \cap D) \leq \mathbb{P}(\max_{i \in B \cap D} Z_i \geq m' + bx) + \mathbb{P}(\max_{i \in S} Z_i \leq m - bx)$$

$$(28) \qquad\qquad \leq e^{-x^2/2} + n^{-\delta^2} + e^{-b^2x^2/2}$$

$$\leq 2e^{-(m-m')^2/8} + n^{-\delta^2}.$$

Thus, by (21), (23), (27), and (28), we have

$$(29) \qquad \mathbb{P}(R(I^0, I^t) \geq r) \leq 6e^{-\delta^2m^2/8} + n^{-\delta^2} \leq 7e^{-\delta^2m^2/8}.$$

Now, if $a \geq \delta$, then

$$
\begin{aligned}
r^2 &= \frac{b^2\zeta - \alpha^2(b^4 - 4b^2\delta + 4\delta^2)}{b^2\zeta} \\
&= \frac{b^2(1 - a^2\alpha^2 + 2a\delta\alpha^2 - \delta^2\alpha^2 + \delta^2) - \alpha^2(b^4 - 4b^2\delta + 4\delta^2)}{b^2\zeta} \\
&\leq \frac{(1 - \alpha^2)b^2 + 2b^2 a\delta\alpha^2 + b^2\delta^2 + 4\alpha^2 b^2\delta}{b^4} \\
&\leq \frac{1 - \alpha^2 + C\delta}{b^2},
\end{aligned}
$$

where C denotes a universal constant (whose value may change from line to line). Again, if $a < \delta$,

$$
\begin{aligned}
r^2 &= \frac{b^2 - \alpha^2(b^4 - 4b^2\delta + 4\delta^2)}{b^2} \\
&\leq 1 - \alpha^2 b^2 + 4\alpha^2\delta \\
&= 1 - \alpha^2 + \alpha^2 a^2 + 4\alpha^2\delta \leq 1 - \alpha^2 + 5\delta.
\end{aligned}
$$

Therefore in all cases we have

$$
\begin{aligned}
\mathbb{E}(R(I^0, I^t)) &\leq r + \mathbb{P}(R(I^0, I^t) \geq r) \\
&\leq \left(\frac{1 - \alpha^2 + C\delta}{b^2}\right)^{1/2} + Ce^{-\delta^2 m^2/8},
\end{aligned}
$$

From the definition (20) of δ we have

$$
e^{-\delta^2 m^2/8} = \frac{2^{1/8}}{m^{1/2}}.
$$

Since $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ and $\alpha \leq 1$, we get

$$
\mathbb{E}(R(I^0, I^t)) \leq \frac{\sqrt{2(1-\alpha)}}{b} + \frac{C(\log m)^{1/4}}{bm^{1/2}}.
$$

Now, $m^{1/2} = \alpha^{1/2}(2\log n)^{1/4}$. Since $R(i, j) \leq 1$ for all i, j, we can put a large enough constant in front of the first term to remove the $\alpha^{1/2}$ from the denominator of the second term. Finally, the bound on v comes from Lemma 3.1 and our bound on $\mathbb{E}(R(I^0, I^t))$. This completes the proof. $\square$

6. Example: Chaotic nature of the first eigenvector

A random Hermitian matrix $A = (a_{ij})_{1 \leq i, j \leq n}$ is said to belong to the Gaussian Unitary Ensemble (GUE) if (i) $(a_{ij})_{1 \leq i \leq j \leq n}$ are independent random variables, (ii) the diagonal entries are standard real Gaussian random variables, and (iii) $(a_{ij})_{1 \leq i < j \leq n}$ are standard complex Gaussian variables (i.e. real and imaginary parts are i.i.d. $N(0, 1/2)$).

Eigenvalues of GUE matrices are among the most widely studied objects in random matrix theory. For a general introduction to the classical random matrix ensembles and results, we refer to the book by Mehta [51]. The study

of the largest eigenvalue was revolutionized through the work of Tracy and
Widom [78, 79, 80]. One of the striking implications of their work is that the
largest eigenvalue has variance of order $n^{-1/3}$, beating the $O(1)$ bound given
by standard isoperimetric and martingale methods. But the Tracy-Widom
result is in the sense of weak convergence, and does not provide an actual
bound on the variance that we need. A variance bound of order $n^{-1/3}$ was
proved by Ledoux [48] and independently by Aubrun [5].

The eigenvectors of GUE matrices, taken as rows (or columns) of a ma-
trix, give rise to a Haar-distributed unitary matrix. In that sense, they
are quite well-understood. However, the behavior of the eigenvectors un-
der perturbations of the matrix has not been studied. Such questions arise,
for instance, in the study of chaos in the spherical Sherrington-Kirkpatrick
model of spin glasses. For the definition of this model and further references,
let us refer to the recent paper of Talagrand and Panchenko [59], where it
was proved that the model is chaotic with respect to an external field. The
goal of this section is to show that the first eigenvector is unstable under
small perturbations of the matrix, and give a quantitative version of this
statement. In the spherical SK model with complex spins, this establishes
chaos with respect to the disorder at zero temperature.

Let us now formulate the question in terms of Gaussian fields. For each
vector $\mathbf{u}$ in the unit sphere S^{2n-1} of $\mathbb{C}^n$, define the quadratic form

$$X_{\mathbf{u}} := \mathbf{u}^* A \mathbf{u} = \sum_{i,j=1}^n a_{ij} \overline{u}_i u_j,$$

where, as usual, $\overline{u}_j$ is the complex conjugate of u_j and $\mathbf{u}^*$ is the adjoint (i.e.
conjugate transposed) of $\mathbf{u}$. Since A is Hermitian, $X_{\mathbf{u}}$ is a real Gaussian
random variable.

Now, if $\mathbf{v} = z\mathbf{u}$ for some z in the unit sphere $U(1)$ of $\mathbb{C}$, then $X_{\mathbf{v}} \equiv X_{\mathbf{u}}$.
Therefore to retain identifiability, we define $X_{\mathbf{u}}$ for each $\mathbf{u}$ not in S^{2n-1}, but
in the complex projective space $\mathbb{CP}^{n-1} = S^{2n-1}/U(1)$. However, we will
continue to write elements of $\mathbb{CP}^{n-1}$ as if they were elements of $\mathbb{C}^n$, with
the quotienting being implicit. With that convention, let

$$\lambda_1 := \max_{\mathbf{u} \in \mathbb{CP}^{n-1}} X_{\mathbf{u}}, \quad \mathbf{u}_1 := \operatorname{argmax}_{\mathbf{u} \in \mathbb{CP}^{n-1}} X_{\mathbf{u}}.$$

Then λ_1 is the largest eigenvalue of the GUE matrix A and $\mathbf{u}_1$ is the corre-
sponding unit eigenvector. Our objective in this section is to show that $\mathbf{u}_1$
is chaotic under small perturbations of A.

Here we must remark that $\mathbf{u}_1$ is almost surely well-defined in $\mathbb{CP}^{n-1}$.
This follows from the fact that the eigenvalues all have multiplicity 1 almost
surely, which can be deduced, for instance, from the well-known joint density
of the eigenvalues of GUE (see e.g. Mehta [51], Chapter 3).

Now let A' be an independent copy of A, and as usual define the perturbed
matrix $A^t := e^{-t}A + \sqrt{1 - e^{-2t}}A'$. Let $\mathbf{u}_1^t$ be the first eigenvector of A^t. We
want to show that $|\langle \mathbf{u}_1, \mathbf{u}_1^t \rangle| := |\sum_{i=1}^n u_{1,i} \overline{u}_{1,i}^t|$ tends to decay rapidly with t.

Note that there is no ambiguity because for any $\mathbf{u}, \mathbf{v} \in \mathbb{CP}^{n-1}$, $|\langle \mathbf{u}, \mathbf{v} \rangle|$ is well-defined (although $\langle \mathbf{u}, \mathbf{v} \rangle$ is not).

Theorem 6.1. *There is a universal constant C such that for any $t \geq 0$,*

$$\mathbb{E}|\langle \mathbf{u}_1, \mathbf{u}_1^t \rangle|^2 \leq \frac{C}{(1 - e^{-t})n^{1/3}},$$

where $\mathbf{u}_1^t$ is the first eigenvector of the perturbed matrix A^t defined above.

Remarks. This shows that whenever $t \gg n^{-1/3}$, the vectors $\mathbf{u}_1$ and $\mathbf{u}_1^t$ are almost orthogonal with high probability. As mentioned before, this result also proves that the ground state of the (complex) spherical Sherrington-Kirkpatrick model is chaotic under small perturbations of the disorder.

Proof. An easy computation gives that for any $\mathbf{u}, \mathbf{v} \in \mathbb{CP}^{n-1}$,

$$\mathrm{Cov}(X_\mathbf{u}, X_\mathbf{v}) = |\langle \mathbf{u}, \mathbf{v} \rangle|^2,$$

where one should note that the right hand side is well-defined on $\mathbb{CP}^{n-1}$. It is known from random matrix theory [48, 5] that $\mathrm{Var}(\lambda_1) \leq Cn^{-1/3}$. Therefore above formula for the covariance and Theorem 3.2 seem to imply that the proof is done. However, we have to be a little careful because Theorem 3.2 works only for Gaussian fields on finite index sets. But this can be easily taken care of by taking the Gaussian field $(X_\mathbf{u})$ restricted to finer and finer nets of points in $\mathbb{CP}^{n-1}$ and using the uniqueness of the maximizer and continuity to pass to the limit. $\qquad\square$

7. Example: Multiple global maxima in the NK fitness landscape

Kauffman and Levin [43] introduced a class of models for the evolution of hereditary systems, which has since become one of the most popular models in evolutionary biology and some other areas. They named it the NK model because there are two parameters, N and K. The model envisions a genome as consisting on N genes, each of which exists as one of two possible alleles. The fitness score of an allele at a given site is determined by the alleles of the K neighboring sites. Other than that, the fitnesses are as simple as possible, namely i.i.d., and the fitnesses of different sites are averaged to get the overall fitness of a genome.

Let us define things formally. The space of all genomes is $\{0, 1\}^N$. Let

$$Y(i; \boldsymbol{\eta}), \ 0 \leq i \leq N - 1, \ \boldsymbol{\eta} \in \{0, 1\}^{K+1}$$

be a collection of i.i.d. random variables, assumed to be standard Gaussian for our purposes. Given $\boldsymbol{\sigma} \in \{0, 1\}^N$, define the 'fitness' of $\boldsymbol{\sigma}$ as

$$F(\boldsymbol{\sigma}) := \sum_{i=0}^{N-1} Y(i; (\sigma_i, \sigma_{i+1}, \ldots, \sigma_{i+K})),$$

where the subscripts of σ are wrapped around, i.e. $\sigma_{N+i} = \sigma_i$. The function $F : \{0, 1\}^N \to \mathbb{R}$ is called the 'fitness landscape', and the main objects of interest are the local and global maxima of this landscape.

The NK model has been extensively studied, but very little of it is rigorous. The first rigorous paper on the model, to the best of our knowledge, was due to Evans and Steinsaltz [24]. These authors used elegant arguments involving max-plus algebras to carry out computations about the global and local maxima of the NK model when K is fixed and $N \to \infty$. One can also find in [24] a beautifully written introduction to the history and motivation behind the model. Further rigorous results were derived using different tools by Durrett and Limic [23] who proved, among other things, a central limit theorem for the maximum fitness when K is fixed and $N \to \infty$. Some results about the local maxima in the case where N and K both tend to infinity were obtained by Limic and Pemantle [49].

Our goal is to show that when N and K both tend to infinity, the fitness landscape has many global near-maxima, which are 'far away' from each other. In other words, there are many nearly globally fittest genomes that are drastically different from each other. To formalize this statement, we first need define what is meant by 'far away' in the space of genomes. The NK model naturally defines the following measure of proximity between two genomes $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$:

$$(30) \qquad p_{N,K}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') := |\{i : (\sigma_i, \ldots, \sigma_{i+K}) = (\sigma'_i, \ldots, \sigma'_{i+K})\}|.$$

This is a natural definition because

$$p_{N,K}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = \mathrm{Cov}(F(\boldsymbol{\sigma}), F(\boldsymbol{\sigma}')).$$

Note also that if K is large, then $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ may be far apart even if the Hamming distance between $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ is relatively small.

To understand the nature of the global maximum, we first have to have an idea about its size. It was shown by Evans and Steinsaltz [24] that the size of global maximum grows linearly in N when K is fixed (in an asymptotic sense). Following their argument, one can further deduce the surprising fact that when divided by N the expected size of the global maximum can be bounded above and below by universal constants that do not depend on K.

Lemma 7.1. *Irrespective of the value of K, we have*

$$\frac{N}{\sqrt{\pi}} \leq \mathbb{E}(\max_{\boldsymbol{\sigma}} F(\boldsymbol{\sigma})) \leq N\sqrt{2 \log 2}.$$

Remark. In fact, the bounds are sharp. The lower bound is achieved when $K = 0$, and the upper bound is achieved (asymptotically) when $K = N - 1$. Moreover, as we will see in the proof, the expected value is an increasing function of K.

Proof. The upper bound is straightforward from Lemma 2.1. For the lower bound, first observe that if $G(\boldsymbol{\sigma})$ is another measure of fitness, corresponding

to the NK model with $K = 0$, then for every $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$, we have

$$\operatorname{Cov}(G(\boldsymbol{\sigma}), G(\boldsymbol{\sigma}')) \geq \operatorname{Cov}(F(\boldsymbol{\sigma}), F(\boldsymbol{\sigma}')),$$

since it is quite clear that $p_{N,K}(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ is a decreasing function of K. Moreover $\operatorname{Var}(F(\boldsymbol{\sigma})) = \operatorname{Var}(G(\boldsymbol{\sigma})) = N$ for every $\boldsymbol{\sigma}$. Therefore by Slepian's lemma,

$$\mathbb{E}(\max_{\boldsymbol{\sigma}} F(\boldsymbol{\sigma})) \geq \mathbb{E}(\max_{\boldsymbol{\sigma}} G(\boldsymbol{\sigma})).$$

Now, if $(Z(i, \eta))_{0 \leq i \leq N-1, \ \eta \in \{0,1\}}$ are the i.i.d. Gaussian fitness scores used to define G, then

$$\max_{\boldsymbol{\sigma}} G(\boldsymbol{\sigma}) = \sum_{i=0}^{N-1} \max\{Z(i, 0), Z(i, 1)\}.$$

It is easy to compute that the expected value of the maximum of two independent standard Gaussian random variables is $\pi^{-1/2}$. This completes the proof. $\qquad\qquad\square$

The following theorem is the main result of this section. It shows that when N and K both tend to infinity, the fitness landscape exhibits the multiple peaks property. We first establish superconcentration using hypercontractivity, and then use Theorem 3.7 to deduce the existence of multiple peaks.

Theorem 7.2. *Suppose $N > K \geq 1$. Then $\operatorname{Var}(\max_{\boldsymbol{\sigma}} F(\boldsymbol{\sigma})) \leq CN/K$, where C is a universal constant. As a consequence, there is another universal constant C' such that for any $a > 1$, with probability at least*

$$1 - C'a^{-1/4}\sqrt{\log a}$$

there exists $A \subseteq \{0, 1\}^N$ such that

    (a) *$|A| \geq a^{1/12}$,*
    (b) *$p_{N,K}(\boldsymbol{\sigma}, \boldsymbol{\sigma}') < a^{-1/4}N$ for all $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in A$, $\boldsymbol{\sigma} \neq \boldsymbol{\sigma}'$, where $p_{N,K}$ is the measure (30) of proximity between genomes, and*
    (c) *for all $\boldsymbol{\sigma} \in A$,*

$$F(\boldsymbol{\sigma}) \geq \max_{\boldsymbol{\sigma}' \in \{0,1\}^N} F(\boldsymbol{\sigma}') - \frac{aN}{K}.$$

Remarks. By Lemma 7.1 and the concentration of $\max F(\boldsymbol{\sigma})$, we know that $\max F(\boldsymbol{\sigma})$ is of order N. This shows that whenever K is large, there are many near-maximal configurations that are 'far apart' in the sense of the proximity measure $p_{N,K}$, since $p_{N,K}$ ranges between 0 and N. The quantification of 'many' and 'far apart' depends on our choice of a. The theorem shows that any large a works, as long as $a \ll K$. Of course, the theorem in its present form does not have any relevance for realistic values of N and K, but we believe that there is room for improvement to an extent that can be of practical significance. Finally, note that the theorem proves strong multiple peaks when K grows faster than $\sqrt{N}$.

Proof. Let $M = \max_{\boldsymbol{\sigma}} F(\boldsymbol{\sigma})$. Let $\hat{\boldsymbol{\sigma}}$ be the maximizing configuration. With obvious notation, we have

$$\frac{\partial M}{\partial Y(i; \boldsymbol{\eta})} = \mathbb{I}_{\{(\hat{\sigma}_i, \ldots, \hat{\sigma}_{i+K}) = (\eta_1, \ldots, \eta_{K+1})\}}.$$

Therefore by Lemma 3.3, Lemma 4.1, and symmetry, we have

$$\begin{aligned}
&\mathrm{Var}(M) \\
&= \int_0^\infty e^{-t} \sum_{(i,\boldsymbol{\eta})} \mathbb{E}\big(\mathbb{I}_{\{(\hat{\sigma}_i, \ldots, \hat{\sigma}_{i+K}) = (\eta_1, \ldots, \eta_{K+1})\}} \mathbb{I}_{\{(\hat{\sigma}_i^t, \ldots, \hat{\sigma}_{i+K}^t) = (\eta_1, \ldots, \eta_{K+1})\}}\big) dt \\
&\leq \int_0^\infty e^{-t} \sum_{(i,\boldsymbol{\eta})} \big[\mathbb{P}((\hat{\sigma}_i, \ldots, \hat{\sigma}_{i+K}) = (\eta_1, \ldots, \eta_{K+1}))\big]^{1+\tanh(t/2)} dt \\
&= \int_0^\infty e^{-t} \sum_{(i,\boldsymbol{\eta})} 2^{-(K+1)(1+\tanh(t/2))} dt \\
&= \int_0^\infty e^{-t} N 2^{-(K+1)\tanh(t/2)} dt \leq \frac{CN}{K},
\end{aligned}$$

where C is a universal constant. Now let $m = \mathbb{E}(M)$ and $v = \mathrm{Var}(M)$. From Lemma 7.1, we know that $N/\sqrt{\pi} \leq m \leq N\sqrt{2\log 2}$, and from the above computation we know that $v \leq CN/K$. Let $\sigma^2 = \max_{\boldsymbol{\sigma}} \mathrm{Var}(F(\boldsymbol{\sigma})) = N$. Take any $a > 1$, and let $l = [a^{1/12}] + 1$, $\epsilon = a^{-1/4} N$, and $\delta = aN/K$. Then

$$\sqrt{\frac{vml^3 \log l}{\delta\epsilon}} + \left(\frac{v\sigma^4 l^5 \log l}{\delta^3 m\epsilon}\right)^{1/4} \leq Ca^{-1/4}\sqrt{\log a} + C\left(\frac{K^2 a^{-7/3} \log a}{N^2}\right)^{1/4},$$

where C is a universal constant. The second term is dominated by the first. The result now follows from Theorem 3.7. $\qquad\square$

8. EXAMPLE: CHAOS IN DIRECTED POLYMERS

We introduced directed polymers in Section 1. Let us now refresh the reader's memory by defining it once again.

The $(1+1)$-dimensional directed polymer model in Gaussian environment is defined as follows. Let $\mathbb{Z}_+^2$ denote the set of all pairs of non-negative integers, with the lattice graph structure. Let $\mathbf{g} = (g_v)_{v \in \mathbb{Z}_+^2}$ be a collection of i.i.d. standard Gaussian random variables. A directed polymer of length n is a sequence of n adjacent points, beginning at the origin, such that each successive point is either to the right or above the previous point. Thus, there are a total of 2^{n-1} directed polymers of length n. Let $\mathcal{P}_n$ denote this set. An element of $\mathcal{P}_n$ will generally be denoted by $p = (v_0, \ldots, v_{n-1})$, where $v_0 = 0$. The energy of a polymer $p = (v_0, \ldots, v_{n-1})$ in the environment $\mathbf{g}$ is defined as

$$E(p) := -\sum_{i=0}^{n-1} g_{v_i}.$$

For each $p \in \mathcal{P}_n$, let $X_p := -E(p)$ denote the 'weight' of the path, and let $\mathbf{X} = (X_p)_{p \in \mathcal{P}_n}$. Our object of interest is the polymer with the minimum energy (i.e. maximum weight), usually called the ground state of the system. Suppressing the subscript n, we simply denote the minimum energy path by P and its energy by $-M$, that is,

$$M := \max_{p \in \mathcal{P}_n} X_p = X_P.$$

Note that for any p, p',

$$\mathrm{Cov}(X_p, X_{p'}) = |p \cap p'|,$$

where $|p \cap p'|$ denotes the number of vertices common to p and p'.

If $\mathbf{g}'$ is an independent copy of $\mathbf{g}$, and we define the perturbed environment $\mathbf{g}^t := e^{-t}\mathbf{g} + \sqrt{1 - e^{-2t}}\mathbf{g}'$, then the energies of the paths defined in the environment $\mathbf{g}^t$ correspond to our usual definition of $\mathbf{X}^t$. Let P^t denote the maximum weight path in the environment $\mathbf{g}^t$ (henceforth, simply the 'maximal path'). The main result of this section is the following.

Theorem 8.1. *For some universal constant C, we have*

$$\mathrm{Var}(M) \leq \frac{Cn}{\log n}.$$

Consequently, for any $t \geq 0$,

$$\mathbb{E}|P \cap P^t| \leq \frac{Cn}{(1 - e^{-t})\log n}.$$

It is easy to see how Theorem 1.1 follows from this result. Another remark is that the superconcentration of the ground state energy has implications about the fluctuations of the polymer shape and specially the end point of the minimum energy polymer; see Wehr and Aizenman [82] for further insights.

One may object that the perturbation described above is not a small perturbation at all, because we are perturbing all coordinates, and moreover independently. Indeed, the regular notion of perturbation in noise-sensitivity theory involves choosing a small fraction of coordinates at random and replacing the weights with independent copies. However, the two notions lead to the same results in practice, because in one case we are giving large perturbations to a small collection of coordinates, while in the other case (i.e. our case) we are giving small perturbations to all coordinates. One can verify that with correct calibration, the two perturbations have similar effects on almost every conceivable summary statistic.

Secondly, our definition of perturbation of a Gaussian field is the one favored by the physicists, who see it as an Ornstein-Uhlenbeck flow over a small period of time.

The chaos property for directed polymers was first argued heuristically by Huse, Henley, and Fisher [38]. It was subsequently studied numerically and theoretically by Zhang [83] and Mézard [52]. A detailed theoretical

study was done by Fisher and Huse [25] and later, another one by Hwa and Fisher [39]. For a more recent work, see da Silveira and Bouchaud [66]. However, none of these papers give rigorous proofs, or even proofs that can be made rigorous with existing technology. To the best of our knowledge, no such proof exists.

The scheme of the proof of Theorem 8.1 is straightforward given the tools we have. We apply Talagrand's technique of bounding variances using hypercontractivity, and use the resulting superconcentration bound to infer chaos via Theorem 1.8 (and the actual bound via Theorem 3.2). However, on our way to applying hypercontractivity, we run into the difficulty that the probability of the optimal path passing through a given vertex is completely unknown; even a reasonable upper bound is unknown. This problem is overcome by a brilliant trick from a paper of Benjamini, Kalai, and Schramm [7], where a similar variance-bounding problem for first passage percolation was tackled. But serious technical issues remain even after this, because adapting the BKS trick requires a lot more effort in this case compared to the case handled in [7], because of the restriction that the paths have to be directed. In fact, this is the sole reason why the proof in its present form does not generalize to dimensions higher than 2, even though the BKS result for first passage percolation holds in all dimensions.

We should mention here that there exists a large amount of rigorous mathematics on directed polymers, even if there is none on the chaos problem. The rigorous study was probably initiated by Imbrie and Spencer [40], and subsequently carried forward by many authors (e.g. [10],[67], [3], [69], [44], [14], [15], [21], [20]). A very notable contribution was due to Johansson [41], who found a miraculous way to do exact computations in the model with Geometric vertex weights instead of Gaussian, showing that M has fluctuations of order $n^{1/3}$, and moreover a Tracy-Widom limiting distribution upon proper centering and scaling. The result was extended to binary edge weights in Johansson [42], Section 5.1 (see also [30]). The recent work of Cator and Groeneboom [16] implies a probabilistic proof of the $n^{1/3}$ fluctuations, but again for Geometric vertex weights.

We need some preparation before embarking on the proof of Theorem 8.1. First of all, let us remind the reader of our convention that C denotes any positive universal constant whose value may change from line to line. This convention will be repeatedly invoked in the remainder of this section. Next, for each $w \in \mathbb{Z}_+^2$, define the translated environment $\mathbf{g}_w = (g_{w,v})_{v \in \mathbb{Z}_+^2}$ as $g_{w,v} := g_{w+v}$. Now fix n, and define $\mathbf{X}_w$, $\mathbf{X}_w^t$, P_w, P_w^t, and M_w as the analogs of $\mathbf{X}$, $\mathbf{X}^t$, P, P^t, and M for the environment $\mathbf{g}_w$. Next, let

(31) $$B := \{(i,j) : 1 \le i, j \le [n^{1/4}]\},$$

and let

$$\overline{M} := \frac{1}{|B|} \sum_{w \in B} M_w.$$

Our first lemma is the following.

Lemma 8.2. *We have*
$$\mathrm{Var}(\overline{M}) \le \frac{Cn}{\log n}.$$

Proof. With obvious notation, we have
$$\frac{\partial \overline{M}}{\partial g_v} = \frac{1}{|B|} \sum_{w \in B} \frac{\partial M_w}{\partial g_v} = \frac{1}{|B|} \sum_{w \in B} \mathbb{I}_{\{v-w \in P_w\}}$$

Therefore by Lemma 3.3 and Lemma 4.1 we have
$$\mathrm{Var}(\overline{M}) = \int_0^\infty e^{-t} \sum_{v \in \mathbb{Z}_+^2} \mathbb{E}\left[\left(\frac{1}{|B|} \sum_{w \in B} \mathbb{I}_{\{v-w \in P_w\}}\right)\left(\frac{1}{|B|} \sum_{w \in B} \mathbb{I}_{\{v-w \in P_w^t\}}\right)\right] dt$$
$$\le \int_0^\infty e^{-t} \sum_{v \in \mathbb{Z}_+^2} \left[\mathbb{E}\left(\frac{1}{|B|} \sum_{w \in B} \mathbb{I}_{\{v-w \in P_w\}}\right)\right]^{1+\tanh(t/2)} dt.$$

Now fix any $v \in \mathbb{Z}_+^2$. Then
$$\mathbb{E}\left(\frac{1}{|B|} \sum_{w \in B} \mathbb{I}_{\{v-w \in P_w\}}\right) = \frac{1}{|B|} \sum_{w \in B} \mathbb{P}(v - w \in P_w)$$
$$= \frac{1}{|B|} \sum_{w \in B} \mathbb{P}(v - w \in P) = \mathbb{E}\left(\frac{1}{|B|} \sum_{w \in B} \mathbb{I}_{\{v-w \in P\}}\right).$$

Since P is an directed path and B is an $[n^{1/4}] \times [n^{1/4}]$ square, we certainly have
$$\sum_{w \in B} \mathbb{I}_{\{v-w \in P\}} \le 2[n^{1/4}] - 1.$$

Combining the last three observations, we have that
$$\mathrm{Var}(\overline{M}) \le \int_0^\infty Ce^{-t} n^{-\frac{1}{4}\tanh(t/2)} \mathbb{E}\left(\frac{1}{|B|} \sum_{v \in \mathbb{Z}_+^2} \sum_{w \in B} \mathbb{I}_{\{v-w \in P\}}\right) dt$$
$$= \int_0^\infty Ce^{-t} n^{-\frac{1}{4}\tanh(t/2)} \mathbb{E}\left(\frac{1}{|B|} \sum_{w \in B} \sum_{v \in \mathbb{Z}_+^2} \mathbb{I}_{\{v-w \in P\}}\right) dt$$
$$= \int_0^\infty Ce^{-t} n^{1-\frac{1}{4}\tanh(t/2)} dt \le \frac{Cn}{\log n}.$$

This completes the proof of the lemma. $\square$

Let us now introduce some further notation. For any $v \in \mathbb{Z}_+^2$, let $|v|$ denote the sum of its coordinates. For any $u, v \in \mathbb{Z}_+^2$, we write $u \le v$ if $v - u \in \mathbb{Z}_+^2$. For any such u, v, let
$$M_{u \to v} := \max\left\{\sum_{u \in p \setminus \{v\}} g_u : p \text{ is a directed path from } u \text{ to } v\right\}.$$

For any $u \in \mathbb{Z}_+^2$, $n \geq 1$, let

$$M_u^n := \max\left\{\sum_{u \in p} g_u : \ p \text{ is a directed path of length } n, \text{ starting at } u\right\}.$$

Let $P_{u \to v}$ and P_u^n denote the maximizing paths.

Clearly, the distribution of M_u^n depends only on n, and the distribution of $M_{u \to v}$ depends only on $v - u$. Let us now prove some inequalities for the expected values of these quantities. In the following 0 can either denote the real number 0, or the point $(0,0)$ in $\mathbb{Z}^2$, depending on the context.

Lemma 8.3. *For any $1 \leq m \leq n$, we have*

$$\mathbb{E}(M_0^n) \geq \mathbb{E}(M_0^m) + C(n - m).$$

Proof. Clearly, it suffices to prove when $n = m+1$. Let $(v_0, \ldots, v_{m-1})$ denote the maximal path of length m. Let u and w be the vertices immediately to the right and above v_{m-1}. Since the definition of the maximal path of length m does not involve vertices v with $|v| \geq m$, it follows that (g_u, g_w) is still a pair of i.i.d. standard Gaussian random variables, even though u and w are random. To complete the proof, note that $M_0^{m+1} \geq M_0^m + \max\{g_u, g_w\}$. $\square$

Lemma 8.4. *For any $v = (v^x, v^y) \in \mathbb{Z}_+^2$, we have*

$$\mathbb{E}(M_{0 \to v}) \leq C|v|\sqrt{H\left(\frac{v^x}{v^x + v^y}\right)},$$

where $H : [0,1] \to \mathbb{R}$ is the function $H(a) = -a \log a - (1 - a) \log(1 - a)$.

Proof. Note that the total number of directed paths from 0 to v is

$$\binom{v^x + v^y}{v^x}.$$

For any $a \in [0, 1]$, we have

$$1 \geq \binom{v^x + v^y}{v^x} a^{v^x}(1 - a)^{v^y}.$$

Taking $a = v^x/(v^x + v^y)$, we see that

$$\log \binom{v^x + v^y}{v^x} \leq -v^x \log a - v^y \log(1 - a) = H(a)|v|.$$

The result now follows by Lemma 2.1. $\square$

Lemma 8.5. *Let P denote the maximal path of length n. Let $v = (v^x, v^y)$ be a vertex with $|v| \leq n - 1$. There is a universal constant $c \in (0, 1)$ such that if $\min\{v^x, v^y\} \leq c|v|$, then*

$$\mathbb{P}(v \in P) \leq 2e^{-c|v|^2/n}.$$

Proof. Let $\widetilde{P}$ denote the directed path of length n, originating from 0 and passing through v, that maximizes the sum of weights among all such paths. Note that this is just the concatenation of the paths $P_{0 \to v}$ and $P_v^{n-|v|}$. Let $\widetilde{M} = \sum_{v \in \widetilde{P}} g_v$. Then note that by Lemma 8.3,

$$\mathbb{E}(\widetilde{M}) = \mathbb{E}(M_{0 \to v}) + \mathbb{E}(M_v^{n-|v|})$$
$$= \mathbb{E}(M_{0 \to v}) + \mathbb{E}(M_0^{n-|v|})$$
$$\leq \mathbb{E}(M_{0 \to v}) + \mathbb{E}(M) - C|v|,$$

where C is the constant from Lemma 8.3. For the rest of this proof, C will denote this constant. From the above bound, we see that if $\mathbb{E}(M_{0 \to v}) \leq C|v|$, then by Proposition 1.3 we have

$$\mathbb{P}(v \in P) = \mathbb{P}(\widetilde{P} = P) \leq \mathbb{P}(\widetilde{M} = M)$$
$$\leq \mathbb{P}\left(\widetilde{M} \geq \mathbb{E}(\widetilde{M}) + \frac{\mathbb{E}(M) - \mathbb{E}(\widetilde{M})}{2} \right)$$
$$+ \mathbb{P}\left(M \leq \mathbb{E}(M) - \frac{\mathbb{E}(M) - \mathbb{E}(\widetilde{M})}{2} \right)$$
$$\leq 2 \exp\left(-\frac{(\mathbb{E}(M) - \mathbb{E}(\widetilde{M}))^2}{8n} \right)$$
$$\leq 2 \exp\left(-\frac{(C|v| - \mathbb{E}(M_{0 \to v}))^2}{8n} \right).$$

Now, the function H in Lemma 8.4 satisfies $H(a) = H(1-a)$ and

$$\lim_{a \to 0} H(a) = \lim_{a \to 1} H(a) = 0.$$

Therefore, there exists a constant $c \in (0,1)$ such that if $\min\{v^x, v^y\} \leq c|v|$, then $\mathbb{E}(M_{0 \to v}) \leq C|v|/2$, where C is the universal constant in Lemma 8.3. This proves the current lemma by choosing c sufficiently small. $\qquad \square$

Lemma 8.6. *Denote the vertices of the maximal path P by $(v_0, \ldots, v_{n-1})$. Extend the path P to an infinite directed path P' by adding a sequence of points $v_n, v_{n+1}, \ldots$, where each v_i is the point immediately to the right of v_{i-1} if i is odd, and immediately above v_{i-1} if i is even. Take any $w \in \mathbb{Z}_+^2 \setminus \{0\}$ with $|w| \leq c(n-1)$, where c is the constant from Lemma 8.5. Let $L := \min\{i : v_i \geq w\}$. Then for any $l \leq n-1$ such that $|w| \leq cl$, we have*

$$\mathbb{P}(L = l) \leq 2e^{-cl^2/n}.$$

We also have $\mathbb{P}(L \geq n) \leq 2|w|e^{-c(n-1)^2/n}$. As a consequence of these bounds, we have $\mathbb{E}(L) \leq C(|w| + \sqrt{n})$ for some universal constant C.

Proof. Let $k := |w|$. It is easy to see that the definition of L ensures that

(32) $k \leq L \leq n + k - 1.$

Take any integer $l \in [k, n-1]$. Since $|v_L| = L$, we have

$$\mathbb{P}(L = l) = \mathbb{P}(v_L = (w^x, l - w^x) \text{ or } v_L = (l - w^y, w^y))$$
$$\leq \mathbb{P}((w^x, l - w^x) \in P) + \mathbb{P}((l - w^y, w^y) \in P).$$

The proof of the first bound can now be completed by applying Lemma 8.5. For the second, note that again by Lemma 8.5,

$$\mathbb{P}(L \geq n) \leq \sum_{i=0}^{w^x - 1} \mathbb{P}((i, n-1-i) \in P) + \sum_{i=0}^{w^y - 1} \mathbb{P}((n-1-i, i) \in P)$$
$$\leq 2|w|e^{-c(n-1)^2/n}.$$

Together with (32), this completes the proof. $\square$

Lemma 8.7. *Let $\mathcal{R}_{n,k}$ be the set of all directed paths whose end points $u = (u^x, u^y)$ and $v = (v^x, v^y)$ satisfy $|u| \leq 2n$, $|v| \leq 2n$, and*

(33) $$\min\{|u^x - v^x|, |u^y - v^y|\} \leq k.$$

Let

$$R_{n,k} := \max_{p \in \mathcal{R}_{n,k}} \frac{|\sum_{v \in p} g_v|}{\sqrt{|p|}},$$

where $|p|$ denotes the number of vertices in p. Then $\mathbb{E}(R_{n,k}^2) \leq Ck \log n$.

Proof. We wish to use Lemma 2.1. Clearly, for each $p \in \mathcal{R}_{n,k}$,

$$\frac{\sum_{v \in p} g_v}{\sqrt{|p|}} \sim N(0, 1).$$

Let us now find an upper bound on the size of $\mathcal{R}_{n,k}$. The end points of a path can be chosen in at most Cn^2 ways. Given the end points, in view of the restriction (33) and the directed nature of the paths, we can employ the same counting argument as in the proof of Lemma 8.4 to conclude that the number of paths is $\leq (Cn)^k$. Thus, $|\mathcal{R}_{n,k}| \leq n^2(Cn)^k$. The claim now follows by Lemma 2.1. $\square$

Lemma 8.8. *For any $w \in \mathbb{Z}_+^2$, we have*

$$\mathbb{E}(M - M_w)^2 \leq C(|w|^2 + |w|\sqrt{n}) \log n.$$

Proof. As before, let $k = |w|$. Since $\mathbb{E}(M^2)$ is bounded by Cn^2 (easily verified by Lemma 2.1), we can assume without loss of generality that $k \leq c(n-1)$. Let P' and L be defined as in Lemma 8.6. Denote the components of v_i by v_i^x and v_i^y. Then we know that either $v_L^x = w^x$ or $v_L^y = w^y$. Define a directed path $P'' = (u_0, \ldots, u_{n-1})$ as follows. If $v_L^x = w^x$, let $u_0 = w$ and let u_i be the point immediately above u_{i-1} for $1 \leq i \leq L - k$. If $v_L^x > w^x$ and $v_L^y = w^y$, let $u_0 = w$ and let u_i be the point immediately to the right of u_{i-1} for $1 \leq i \leq L - k$. Note that in either case, we end up with $u_{L-k} = v_L$ because $|v_L| = L$. Thereafter, merge the path with P', that is, let

$$u_i = v_{k+i} \quad \text{for} \quad L - k + 1 \leq i \leq n - 1.$$

Let $\mathcal{R}_{n,k}$ and $R_{n,k}$ be defined as in Lemma 8.7 above. Clearly, the paths $(v_0, \ldots, v_{\min\{L, n-1\}})$ and $(u_0, \ldots, u_{L-k})$ belong to $\mathcal{R}_{n,k}$. With this observation, and the fact that P'' is a directed path of length n starting at w, we have

$$
\begin{aligned}
M_w &\geq \sum_{i=0}^{n-1} g_{u_i} = \sum_{i=0}^{L-k} g_{u_i} + \sum_{i=L}^{n+k-1} g_{v_i} \\
&= M - \sum_{i=0}^{\min\{L, n-1\}} g_{v_i} + \sum_{i=\max\{n, L\}}^{n+k-1} g_{v_i} + \sum_{i=0}^{L-k} g_{u_i} \\
&\geq M - 2\sqrt{L} R_{n,k} - k \max_{|v| \leq 2n} |g_v|.
\end{aligned}
$$

Therefore, by the Cauchy-Schwarz inequality, Lemma 2.1, Lemma 8.6, and Lemma 8.7, it follows that

$$
(34) \qquad \mathbb{E}(M - M_w)_+^2 \leq C(|w|^2 + \sqrt{n}|w|) \log n.
$$

Next, let us get a bound on $\mathbb{E}(M_w - M)_+^2$. Denote the vertices of P_w by $(v_{w,0}, \ldots, v_{w,n-1})$. Let $(u_0, \ldots, u_{k-1})$ be any fixed directed path with $u_0 = 0$ and $u_{k-1} = w$. Let P_w' be the directed path

$$
(u_0, \ldots, u_{k-1}, w + v_{w,1}, w + v_{w,2}, \ldots, w + v_{w,n-k}).
$$

Then note that

$$
\begin{aligned}
M &\geq \sum_{v \in P_w'} g_v = M_w + g_{u_0} + \cdots + g_{u_{k-2}} - g_{w+v_{w,n-k+1}} - \cdots - g_{w+v_{w,n-1}} \\
&\geq M_w - 2(k-1) \max_{|v| \leq 2n} |g_v|.
\end{aligned}
$$

This shows that

$$
(35) \qquad \mathbb{E}(M_w - M)_+^2 \leq C|w|^2 \log n.
$$

Combining (34) and (35), our proof is done. $\qquad\square$

Proof of Theorem 8.1. For each $w \in B$, we have

$$
\mathbb{E}(M - M_w)^2 \leq C(|w|^2 + \sqrt{n}|w|) \log n \leq Cn^{3/4} \log n.
$$

Therefore,

$$
\mathbb{E}(M - \overline{M})^2 \leq Cn^{3/4} \log n.
$$

Thus, by Lemmas 8.2 and 8.8 we have

$$
\begin{aligned}
\mathrm{Var}(M) &\leq \mathbb{E}(M - \mathbb{E}(\overline{M}))^2 \\
&\leq 2\mathbb{E}(M - \overline{M})^2 + 2\mathrm{Var}(\overline{M}) \\
&\leq Cn^{3/4} \log n + \frac{Cn}{\log n}.
\end{aligned}
$$

This completes the proof. $\qquad\square$

9. Example: Generalized SK model of spin glasses

Let n be a positive integer and let $\Sigma_n = \{-1, 1\}^n$. Suppose we have n magnetic particles, each having spin $+1$ or -1. A typical element of $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n) \in \Sigma_n$ is called a 'configuration' of spins. Let $g = (g_{ij})_{1 \le i < j \le n}$ be a collection of independent standard Gaussian random variables. Given a realization of g, define the energy of a configuration $\boldsymbol{\sigma}$ as

$$H_n(\boldsymbol{\sigma}) = -\frac{1}{\sqrt{n}} \sum_{1 \le i < j \le n} g_{ij} \sigma_i \sigma_j.$$

The energy landscape so defined corresponds to the famous Sherrington-Kirkpatrick model of spin glasses [65] in the absence of an external field. In the last thirty years, the SK model has been an inspiration for a large body of groundbreaking physics (see the Mézard-Parisi-Virasoro book [53]) as well as beautiful rigorous mathematics (see e.g. [2], [26], [19], [63], [73], [33], [34], [77], [76]). An extensive collection of rigorous results can be found in Talagrand's book [74], a new edition of which is in preparation.

A natural variant of the SK model is the p-spin SK model, where the energy of a configuration is defined as

$$(36) \qquad H_{n,p}(\boldsymbol{\sigma}) = -\frac{1}{n^{(p-1)/2}} \sum_{1 \le i_1, i_2, \dots, i_p \le n} g_{i_1 i_2 \cdots i_p} \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_p},$$

where again, $(g_{i_1 i_2 \cdots i_p})_{1 \le i_1, \dots, i_p \le n}$ is a fixed realization of i.i.d. standard Gaussian random variables. (Usually, the sum is taken over distinct $i_1, \dots, i_p$. We take it over all $i_1, \dots, i_p$ to avoid certain technical inconveniences.) The p-spin model was suggested by Derrida, and subsequently studied by Gross and Mézard [32] and Gardner [27].

A generalized version of the SK model that covers all p-spin models was considered by Talagrand in his celebrated paper on the Parisi formula [76]. It is simply a linear combination of the p-spin energies over all p and covers all cases considered till now. Given a sequence of non-negative real numbers $\mathbf{c} = (c_2, c_3, \dots)$ such that

$$\sum_{p=2}^{\infty} c_p = 1,$$

define the energy function

$$H_{n,\mathbf{c}}(\boldsymbol{\sigma}) := \sum_{p=2}^{\infty} c_p^{1/2} H_{n,p}(\boldsymbol{\sigma}),$$

where $H_{n,p}$ is the p-spin energy defined above in (36). Then the usual SK model corresponds to the sequence $(1, 0, 0, \dots)$, and the p-spin model corresponds to the sequence that has 1 at the pth position and 0 elsewhere.

The objective of this section is to analyze the energy landscape of the generalized SK model. In particular, we are interested in the behavior of

the ground state, i.e. the configuration with minimum energy, and the fluctuations of the energy of the ground state.

There are two notable conjectures about the ground state of the SK model, both seemingly beyond the reach of current technology. One is that the ground state exhibits chaos (exactly in our sense). A physics proof for chaos was given by McKay, Berger, and Kirkpatrick [50]. The second conjecture is about the 'fluctuation exponent' of the ground state energy.

Definition 9.1. *The energy of the ground state is said to have fluctuation exponent ρ if it has fluctuations of order $n^{\rho+o(1)}$.*

Classical results like Proposition 1.2 and Proposition 1.3 imply that for the generalized SK model, $\rho \leq 1/2$. It is predicted by physicists [45, 22] that $\rho = 1/6$ in the SK model, although the prediction has not yet been reliably verified by simulations.

As we know from our Theorem 1.8, the two problems are related. Indeed, superconcentration and hence chaos happens if the fluctuation exponent is anything strictly less than $1/2$. The main result of this section is that the fluctuation exponent of the ground state energy in the generalized SK model is at most $3/8$ if c_p decreases to zero sufficiently slowly as $p \to \infty$.

Theorem 9.2. *Let $I(x) := \frac{1}{2}((1+x)\log(1+x) + (1-x)\log(1-x))$, and $(c_p^*)_{p\geq 2}$ be constants such that for all $x \in (-1, 1)$,*

$$\frac{I(x)}{2\log 2 - I(x)} = \sum_{p=2}^{\infty} c_p^* x^p.$$

Then $c_p^* \geq 0$ for all p and $\sum c_p^* = 1$. Suppose $\mathbf{c} = (c_p)_{p\geq 2}$ is any non-negative sequence such that $\sum c_p = 1$, and for all $r \geq 2$,

$$(37) \qquad \qquad \sum_{p=2}^{r} c_p \leq \sum_{p=2}^{r} c_p^*.$$

Then the ground state energy of the generalized SK model defined by the sequence $\mathbf{c}$ has fluctuation exponent $\leq 3/8$, and consequently, the ground state is chaotic.

Incidentally, it was shown by Wehr and Aizenman [82] that the lattice spin glass (i.e. the Edwards-Anderson model) is not superconcentrated, and hence, not chaotic in our sense. Therefore, if we are looking for chaos in spin glasses, the only option is to look in mean-field models.

The proof of Theorem 9.2 is based on our result about extremal Gaussian fields, namely, Theorem 5.1. The minorizing condition (37) suffices to guarantee extremality of the energy landscape considered as a Gaussian field. We will actually prove a more general version of Theorem 9.2, with precise quantitative bounds. Fix n, and consider a centered Gaussian field

$\mathbf{X} = (X_{\boldsymbol{\sigma}})_{\boldsymbol{\sigma} \in \{-1,1\}^n}$ satisfying

$$\operatorname{Cov}(X_{\boldsymbol{\sigma}}, X_{\boldsymbol{\sigma}'}) = \xi\left(\frac{\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}'}{n}\right) \quad \text{for all } \boldsymbol{\sigma} \in \{-1,1\}^n,$$

where $\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}' = \sum_{i=1}^n \sigma_i \sigma_i'$ is the usual inner product, and $\xi : [-1,1] \to [-1,1]$ is a function with $\xi(1) = 1$. Let $\mathbf{X}'$ be an independent copy of $\mathbf{X}$. As usual, define the perturbed field

$$\mathbf{X}^t = e^{-t}\mathbf{X} + \sqrt{1 - e^{-2t}}\mathbf{X}'.$$

Let $\hat{\boldsymbol{\sigma}}^t = \operatorname{argmax}_{\boldsymbol{\sigma}} X_{\boldsymbol{\sigma}}^t$. The following theorem is a quantitative version of Theorem 9.2.

Theorem 9.3. *Let $I(x)$ be as in Theorem 9.2. Suppose that*

(38) $\quad |\xi(x)| \le \xi(|x|) \quad and \quad \xi(x) \le \dfrac{I(x)}{2\log 2 - I(x)} \quad for\ all\ \ x \in (-1,1).$

Then we have the bounds

$$\operatorname{Var}(\max_{\boldsymbol{\sigma}} X_{\boldsymbol{\sigma}}) \le C\left(\frac{\log n}{n}\right)^{1/4}, \quad and$$

$$\mathbb{E}\left(\xi\left(\frac{\hat{\boldsymbol{\sigma}}^0 \cdot \hat{\boldsymbol{\sigma}}^t}{n}\right)\right) \le \frac{C}{\sqrt{1 - e^{-2t}}}\left(\frac{\log n}{n}\right)^{1/4},$$

where C is a universal constant.

Before proving Theorem 9.3, let us first show that it implies Theorem 9.2. First of all, if we define $X_{\boldsymbol{\sigma}} = -n^{-1/2}H_{n,\mathbf{c}}(\boldsymbol{\sigma})$, then

$$\operatorname{Cov}(X_{\boldsymbol{\sigma}}, X_{\boldsymbol{\sigma}'}) = \sum_{p=2}^{\infty} c_p\left(\frac{\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}'}{n}\right)^p.$$

Thus, we are in the setting of Theorem 9.3 with $\xi(x) = \sum c_p x^p$. So if we can show that (38) follows from (37), then Theorem 9.3 would imply that $\operatorname{Var}(\max_{\boldsymbol{\sigma}} X_{\boldsymbol{\sigma}}) \le n^{-1/4+o(1)}$ and hence that

$$\operatorname{Var}(\min_{\boldsymbol{\sigma}} H_{n,\mathbf{c}}(\boldsymbol{\sigma})) \le n^{3/4+o(1)},$$

which proves the claim. The implication of (38) from (37) is proved as follows. First, it is easy to verify that the power series for $I(x)$ has non-negative coefficients, and therefore so does

$$\frac{I(x)}{2\log 2 - I(x)} = \sum_{k=1}^{\infty}\left(\frac{I(x)}{2\log 2}\right)^k.$$

For each r, let $C_r = \sum_{p=2}^r c_p$, and $C_r^* = \sum_{p=2}^r c_p^*$, with $C_1 = C_1^* = 0$. The assumption (37) says that $C_r \le C_r^*$ for each r. Thus for any $x \in (0, 1)$,

$$
\begin{aligned}
\xi(x) = \sum_{p=2}^\infty c_p x^p &= \sum_{p=2}^\infty (C_p - C_{p-1}) x^p \\
&= \sum_{p=2}^\infty (x^p - x^{p+1}) C_p \\
&\le \sum_{p=2}^\infty (x^p - x^{p+1}) C_p^* = \sum_{p=2}^\infty c_p^* x^p = \frac{I(x)}{2\log 2 - I(x)}.
\end{aligned}
$$

Since $|\xi(x)| \le \xi(|x|)$ and I is symmetric, the inequality holds for $x \in (-1, 0]$ as well. This completes the argument for Theorem 9.2.

Proof of Theorem 9.3. We use Theorem 1.11. Let $\mathbf{1}$ denote the configuration of all 1's. Then by symmetry, we have

$$
\sum_{\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \{-1,1\}^n} 2^{-2n/(1+\xi(\frac{\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}'}{n}))} = 2^n \sum_{\boldsymbol{\sigma} \in \{-1,1\}^n} \exp\left(-\frac{2n\log 2}{1 + \xi(\frac{\mathbf{1} \cdot \boldsymbol{\sigma}}{n})}\right).
$$

By the binomial theorem, we know that the number of configurations that have $\sum_{i=1}^n \sigma_i = k$ is exactly

$$
\binom{n}{\frac{n+k}{2}},
$$

which is interpreted as zero if k and n have different parity. Now, we have that for any $p \in [0, 1]$,

$$
\binom{n}{\frac{n+k}{2}} p^{(n+k)/2}(1-p)^{(n-k)/2} \le 1.
$$

Taking $p = (n + k)/2n$, we get

$$
\binom{n}{\frac{n+k}{2}} \le e^{n(\log 2 - I(k/n))},
$$

where $I(x)$ is defined in the statement of the theorem. Again, the hypothesis implies that

$$
\frac{2\log 2}{1 + \xi(x)} \ge 2\log 2 - I(x) \quad \text{for all} \quad x \in [-1, 1],
$$

Thus, we have

$$
\sum_{\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \{-1,1\}^n} 2^{-2n/(1+\xi(\frac{\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}'}{n}))} \le 2^n \sum_{k=-n}^n \exp\left(-\frac{2n\log 2}{1 + \xi(k/n)}\right)\binom{n}{\frac{n+k}{2}}
$$

$$
\le Cn.
$$

The proof now follows from Theorem 1.11. $\square$

10. Example: The Discrete Gaussian Free Field

In this section, we show that the Discrete Gaussian Free Field (DGFF) on an $n \times n$ grid (defined below) is a superconcentrated Gaussian field. The massless Gaussian free field is an important mathematical object, inspiring a substantial amount of rigorous literature. It is essentially a higher dimensional analog of Brownian motion, where the dimension of time (rather than space) is higher than one. Although initially introduced as a toy model for the Ising interface, the topic has grown in its own right and has found important intersections with subjects as diverse as quantum gravity and stochastic Loewner evolutions. The DGFF is a finite approximation to the massless free field, just as random walk is a finite approximation of Brownian motion. For further motivation, definitions, and a review of the rigorous literature, we refer to Giacomin [28] and the excellent recent survey of Sheffield [64].

10.1. **Zero boundary condition.** Let $V_n := \{0, \ldots, n-1\}^2$, and ∂V_n be the inner boundary, that is, the points in V_n which have a nearest neighbor outside. Let $\mathrm{int}(V_n) := V_n \backslash \partial V_n$. The two-dimensional discrete Gaussian free field on V_n with zero boundary condition is defined as a family $\Phi_n = \{\phi_x\}_{x \in V_n}$ of centered gaussian random variables with covariances given by the discrete Green's function of the (discrete) Laplacian on $\mathrm{int}(V_n)$. This means, explicitly, that $\phi_x \equiv 0$ for $x \in \partial V_n$, and

$$(39) \qquad \mathrm{Cov}(\phi_x, \phi_y) = G_n(x, y) = \mathbb{E}_x \left(\sum_{i=0}^{\tau_{\partial V_n}} \mathbb{I}_{\{\eta_i = y\}} \right), \quad x, y \in \mathrm{int}(V_n),$$

where $\{\eta_i\}_{i \geq 0}$ is a standard symmetric nearest neighbor random walk on the two-dimensional lattice $\mathbb{Z}^2$, starting at x, and $\tau_{\partial V_n}$ is the first entrance time in ∂V_n. The law of Φ_n is the Gaussian distribution with density function proportional to

$$(40) \qquad \exp\left(-\frac{1}{8} \sum_{x \sim y} (\phi_x - \phi_y)^2 \right),$$

where $x \sim y$ means x and y are neighbors in V_n (each pair counted once), with the understanding that we set $\phi_x \equiv 0$ for $x \in \partial V_n$ in the above formula. This can be seen as follows. Fix $y \in \mathrm{int}(V_n)$, and for each $x \in \mathrm{int}(V_n)$ let $f(x)$ and $g(x)$ denote the left and right sides of (39). Using (40) it is easy to show that for any x,

$$\mathbb{E}(\phi_x | (\phi_z)_{z \neq x}) = \frac{1}{4} \sum_{z \in V_n, \, z \sim x} \phi_z.$$

It follows that the function f is discrete harmonic on $\mathrm{int}(V_n) \backslash \{y\}$, that is,

$$f(x) = \mathbb{E}(\mathbb{E}(\phi_x | (\phi_z)_{z \neq x}) \phi_y) = \frac{1}{4} \sum_{z \in V_n, \, z \sim x} f(z).$$

Similarly, it is easy to show that $g(x)$ is discrete harmonic on $\text{int}(V_n)\backslash\{y\}$ by conditioning on η_1. Again, putting

$$\bar{\phi}_y = \mathbb{E}(\phi_y | (\phi_z)_{z \neq y}) = \frac{1}{4} \sum_{z \in V_n, \ z \sim y} \phi_z,$$

we have

$$f(y) - \frac{1}{4} \sum_{z \in V_n, \ z \sim y} f(z) = \mathbb{E}((\phi_y - \bar{\phi}_y)\phi_y)$$

$$= \mathbb{E}((\phi_y - \bar{\phi}_y)^2) + \mathbb{E}((\phi_y - \bar{\phi}_y)\bar{\phi}_y)$$

$$= \mathbb{E}(\text{Var}(\phi_y | (\phi_z)_{z \neq y})) + 0$$

$$= 1.$$

Similarly, conditioning on η_1, we can show

$$g(y) - \frac{1}{4} \sum_{z \in V_n, \ z \sim y} g(z) = 1.$$

Thus, $f - g$ is discrete harmonic on $\text{int}(V_n)$. But $f = g = 0$ on ∂V_n. Thus we must have $f = g$ everywhere on V_n, which proves that the density (40) indeed corresponds to the Gaussian field with covariance (39).

It was shown by Bolthausen, Deuschel, and Giacomin ([11], Lemma 1) that as $n \to \infty$, we have

$$\max_{y \in V_n} \text{Var}(\phi_y) = \frac{2}{\pi} \log n + O(1).$$

The unexpected and surprising fact, also in the same paper ([11], Theorem 2), is that as $n \to \infty$,

$$\mathbb{E}(\max_{y \in V_n} \phi_y) \sim 2\sqrt{\frac{2}{\pi} \log n},$$

exactly as if $\{\phi_y\}_{y \in V_n}$ were independent. Here $a_n \sim b_n$ means, as usual, that $\lim_{n \to \infty} a_n/b_n = 1$. In our terminology, the DGFF with zero boundary condition is extremal. Combined with Theorem 5.1, a direct consequence is the following result.

Proposition 10.1. *The discrete Gaussian free field on an $n \times n$ grid with zero boundary condition is superconcentrated (as $n \to \infty$), meaning that*

$$\text{Var}(\max_{y \in V_n} \phi_y) = o(\log n).$$

Consequently, the two dimensional DGFF is chaotic and has the multiple peaks property.

An immediate comment is that we cannot give a bound on the variance better than $o(\log n)$. This is because the result about the asymptotic behavior of $\mathbb{E}(\max \phi_y)$ available from [11] does not include any explicit rate of convergence, which makes it impossible for us to get a more informative bound from Theorem 5.1.

We next consider the free field on an $n \times n$ torus, where we can take advantage of the symmetry to prove that $\mathrm{Var}(\max \phi(y)) = O(\log \log n)$.

10.2. **DGFF on a torus.** Let $\mathbb{T}_n$ be the set $\{0, \ldots, n-1\}^2$ endowed with the graph structure of a torus, that it, (a, b) and (c, d) are adjacent if $a - c \equiv \pm 1 \pmod{n}$ and $b - d \equiv \pm 1 \pmod{n}$. We wish to define a Gaussian free field on this graph. However, the graph has no natural boundary. The easiest (and perhaps the most natural) way to overcome this problem is to modify the definition (39) of the covariance by replacing the stopping time $\tau_{\partial V_n}$ with a random time τ that is of the same order of magnitude as $\tau_{\partial V_n}$, but is independent of all else. Specifically, we prescribe

$$(41) \qquad \mathrm{Cov}(\phi_x, \phi_y) = \mathbb{E}_x \left(\sum_{i=0}^{\tau} \mathbb{I}_{\{\eta_i = y\}} \right),$$

where $(\eta_i)_{i \geq 0}$ is a simple random walk on the torus started at x, and τ is a random variable independent of $(\eta_i)_{i \geq 0}$, which we take to be Geometrically distributed with mean n^2. The reason for the particular choice of the Geometric distribution is that it translates into a simple modification of the density (40) by the introduction of a small 'mass'; the new density function turns out to be

$$(42) \qquad \exp\left(-\frac{(1 - \frac{1}{n^2})}{8} \sum_{x \sim y} (\phi_x - \phi_y)^2 - \frac{1}{2n^2} \sum_x \phi_x^2 \right).$$

Here $x \sim y$ means that x and y are neighbors on the torus. To show that this density indeed corresponds to that of a centered Gaussian field with covariance (41), we proceed as in the case of the DGFF with zero boundary condition. Fix $y \in \mathbb{T}_n$, and let $f(x)$ and $g(x)$ be the two sides of (41). From (41) and (42), one can check using conditional expectations that for $x \neq y$,

$$f(x) = \frac{n^2 - 1}{4n^2} \sum_{z \in V_n, \, z \sim x} f(z), \quad g(x) = \frac{n^2 - 1}{4n^2} \sum_{z \in V_n, \, z \sim x} g(z).$$

(The second identity holds because conditional on $\tau \geq 1$, $\tau - 1$ has the same distribution as τ. This is where we use that τ has a Geometric law.) Again, using similar computations as before, it can be checked that

$$f(y) = 1 + \frac{n^2 - 1}{4n^2} \sum_{z \in V_n, \, z \sim x} f(z),$$

$$g(y) = 1 + \frac{n^2 - 1}{4n^2} \sum_{z \in V_n, \, z \sim y} g(z).$$

Combining the last two displays, one can conclude that $|f - g|$ is a nonnegative strictly subharmonic function on $\mathbb{T}_n$, which implies that it must be zero everywhere.

Although we assign a small mass in our definition of the DGFF on the torus, we can still call it a 'massless free field' in an asymptotic sense because

the stopping times $\tau_{\partial V_n}$ and τ are both of order n^2 and it is not difficult to show that the covariances in the two models differ by $O(1)$.

Let us now state the main result of this section, which shows that the DGFF on the torus is superconcentrated, with an explicit bound of order $\log \log n$ on the variance of the maximum.

Theorem 10.2. *Let $(\phi_x)_{x \in \mathbb{T}_n}$ be the DGFF on the torus defined above. For some universal constant C, we have*

$$\mathrm{Var}(\max_{x \in \mathbb{T}_n} \phi_x) \leq C \log \log n.$$

The proof of this result is via the use of hypercontractivity, more specifically Theorem 4.3. The key advantage in the torus model is that we know that the location of the maximum is uniformly distributed. In the zero boundary situation, we had very little information about the location of the maximum.

Let us now proceed to prove Theorem 10.2. The first step is a basic observation about the simple symmetric random walk on $\mathbb{Z}$.

Lemma 10.3. *Suppose $(\alpha_i)_{i \geq 0}$ is a simple symmetric random walk on $\mathbb{Z}$, starting at 0. Then for any $k \in \mathbb{Z}$ and $i \geq 1$, we have*

$$\mathbb{P}(\alpha_i = k) \leq \frac{Ce^{-k^2/4i}}{\sqrt{i}}$$

where C is a universal constant.

Proof. If $|k| > i$ or $k \not\equiv i \pmod 2$, then $\mathbb{P}(\alpha_i = k) = 0$ and we have nothing to prove. If $|k| = i$, we have $\mathbb{P}(\alpha_i = k) = 2^{-i}$, which is consistent with the statement of the lemma. In all other cases,

$$\mathbb{P}(\alpha_i = k) = \binom{i}{\frac{i+k}{2}} 2^{-i}.$$

Using the Stirling approximation, we get

$$\mathbb{P}(\alpha_i = k) \leq \frac{C}{\sqrt{i}} \exp\left(-\frac{i+k+1}{2} \log\left(1 + \frac{k}{i}\right) - \frac{i-k+1}{2} \log\left(1 - \frac{k}{i}\right)\right)$$

$$= \frac{C}{\sqrt{i}} \exp\left(-iI\left(\frac{k}{i}\right) - \frac{1}{2} \log\left(1 - \frac{k^2}{i^2}\right)\right),$$

where

$$I(x) = \frac{1+x}{2} \log(1+x) + \frac{1-x}{2} \log(1-x).$$

The function I has a power series expansion

$$I(x) = \sum_{p=1}^{\infty} \frac{x^{2p}}{(2p-1)2p}, \quad x \in (-1, 1).$$

In particular, $I(x) \geq x^2/2$. This inequality suffices to prove the lemma when, say, $k \leq i/2$. On the other hand, if $i/2 < k < i$ and i is so large that

$\log i \le i/8$ (which can be assumed by choosing a suitably large C), we have

$$-\frac{1}{2} \log\left(1 - \frac{k^2}{i^2}\right) \le \frac{\log i}{2} \le \frac{i}{16} \le \frac{k^2}{4i},$$

which implies that

$$\mathbb{P}(\alpha_i = k) \le \frac{2}{\sqrt{2\pi i}} \exp\left(-\frac{k^2}{2i} + \frac{k^2}{4i}\right).$$

This completes the proof. $\square$

We are going to use random walks on $\mathbb{Z}^2$ to produce random walks on $\mathbb{T}_n$. For this purpose, we observe that there is a natural map $Q_n : \mathbb{Z}^2 \to \mathbb{T}_n$ which takes a point $(x_1, x_2) \in \mathbb{Z}^2$ to the unique point (x_1', x_2') in $\mathbb{T}_n$ satisfying $x_1 \equiv x_1' \pmod{n}$ and $x_2 \equiv x_2' \pmod{n}$. For any $x, y \in \mathbb{T}_n$, define the toric Euclidean distance $d_n(x, y)$ as:

$$d_n(x, y) := d(x, Q_n^{-1}(y)),$$

where $d(x, A)$ is the usual Euclidean distance of a point x from a set A. It is not difficult to verify that actually $d_n(x, y) = d(Q_n^{-1}(x), Q_n^{-1}(y))$ and therefore the definition is symmetric in x and y.

Lemma 10.4. *Let $(\eta_i)_{i \ge 0}$ be a simple symmetric random walk on $\mathbb{T}_n$. Then for any $x, y \in \mathbb{T}_n$ and any $i \ge 1$, we have*

$$\mathbb{P}_x(\eta_i = y) \le \begin{cases} C i^{-1} e^{-d_n(x,y)^2/4i} & \text{if } i \le n^2, \\ C n^{-2} & \text{if } i > n^2, \end{cases}$$

where C is a universal constant and $\mathbb{P}_x$ denotes the law of the random walk starting from x.

Proof. Let $(\beta_i)_{i \ge 0}$ be a simple symmetric random walk on $\mathbb{Z}^2$. Then a random walk on the torus is easily obtained as $\eta_i = Q_n(\beta_i)$. For any $x, y \in \mathbb{T}_n$, we have

$$(43) \qquad \mathbb{P}_x(\eta_i = y) = \sum_{z \in Q_n^{-1}(y)} \mathbb{P}_x(\beta_i = z).$$

Now, for any x and z, Lemma 10.3 shows that

$$(44) \qquad \mathbb{P}_x(\beta_i = z) \le \frac{C}{i} \exp\left(-\frac{d(x,z)^2}{4i}\right),$$

where $d(x, z)$ is the Euclidean distance between x and z. Now fix $x, y \in \mathbb{T}_n$, and let z be the nearest point to x in $Q_n^{-1}(y)$. Then, if $x = (x_1, x_2)$ and $z = (z_1, z_2)$, we have by (43) and (44) that

$$\mathbb{P}_x(\eta_i = y) \le \frac{C}{i} \sum_{k_1, k_2 \in \mathbb{Z}} \exp\left(-\frac{(x_1 - z_1 + k_1 n)^2 + (x_2 - z_2 + k_2 n)^2}{4i}\right).$$

It is easy to see that $|x_j - z_j| \le n/2$ for $j = 1, 2$ (otherwise, we can choose a 'better' z_j so that z is closer to x.) Thus, for $j = 1, 2$,

$$(x_j - z_j + k_j n)^2 = (x_j - z_j)^2 + k_j^2 n^2 + 2(x_j - z_j)k_j n$$
$$\ge (x_j - z_j)^2 + |k_j|(|k_j| - 1)n^2.$$

Therefore,

$$\mathbb{P}_x(\eta_i = y) \le \frac{C}{i} e^{-d(x,z)^2/4i} \sum_{k_1, k_2 = 0}^{\infty} \exp\left(-\frac{(k_1(k_1 - 1) + k_2(k_2 - 1))n^2}{4i}\right)$$
$$\le \frac{C}{i} e^{-d(x,z)^2/4i} \left(1 + \sum_{k=1}^{\infty} e^{-k^2 n^2/4i} + \sum_{r,s=1}^{\infty} e^{-(r^2+s^2)n^2/4i}\right).$$

Comparing the last two terms with integrals, we get

$$\mathbb{P}_x(\eta_i = y) \le \frac{C}{i} e^{-d(x,z)^2/4i} \left(1 + \int_{\mathbb{R}} e^{-u^2 n^2/4i} du + \int_{\mathbb{R}^2} e^{-(u^2+v^2)n^2/4i} \, du \, dv\right)$$
$$\le \frac{C}{i} e^{-d(x,z)^2/4i} \left(1 + \frac{\sqrt{i}}{n} + \frac{i}{n^2}\right).$$

It is not difficult to verify by considering the cases $i \le n^2$ and $i > n^2$ that this completes the proof. $\qquad\square$

Lemma 10.5. *For any $x \ne y \in \mathbb{T}_n$, we have*

$$0 \le \mathrm{Cov}(\phi_x, \phi_y) \le C \log \frac{n}{d_n(x, y)} + C.$$

and $\mathrm{Var}(\phi_x) \le C \log n$, where C is a universal constant.

Proof. From the representation (41), we see that the covariances are nonnegative. Now fix two distinct points $x, y \in \mathbb{T}_n$, and let $d = d_n(x, y)$. From (41), we have

$$\mathrm{Cov}(\phi_x, \phi_y) = \sum_{i=0}^{\infty} \mathbb{P}_x(\eta_i = y)\mathbb{P}(\tau \ge i)$$
$$\le \sum_{i=1}^{n^2} \frac{Ce^{-d^2/4i}}{i} + \sum_{i=n^2}^{\infty} \frac{C}{n^2}\left(1 - \frac{1}{n^2}\right)^i.$$

Clearly, the second sum can be bounded by a constant that does not depend on n. For the first, note that by the inequality $e^{-x} \le x^{-1}$ that holds for $x \ge 1$, we have

$$\sum_{1 \le i \le n^2} \frac{Ce^{-d^2/4i}}{i} = \sum_{1 \le i \le d^2} \frac{Ce^{-d^2/4i}}{i} + \sum_{d^2 < i \le n^2} \frac{Ce^{-d^2/4i}}{i}$$
$$\le \sum_{1 \le i \le d^2} \frac{C}{d^2} + \sum_{d^2 < i \le n^2} \frac{C}{i}$$
$$\le C + C(\log n^2 - \log d^2).$$

The bound on the variance follows similarly. This completes the proof. $\square$

Proof of Theorem 10.2. Fix some $r \le C \log n$, where C is the same constant as in Lemma 10.5. If $\mathrm{Cov}(\phi_x, \phi_y) \ge r$, then by Lemma 10.5,

$$(45) \qquad d_n(x, y) \le n e^{1-(r/C)} =: s.$$

Let A be an s-net of $\mathbb{T}_n$ for the metric d_n (i.e. a set of points that are mutually separated from each other by distance $> s$, and is maximal with respect to this property). Let $\mathcal{C}(r)$ be the collection of all $2s$-balls around the points of A. Then by (45) we see that whenever $\mathrm{Cov}(\phi_x, \phi_y) \ge r$, we must have that $x, y \in D$ for some $D \in \mathcal{C}(r)$. By symmetry, we see that the probability of the maximum being at any point $z \in \mathbb{T}_n$ is exactly n^{-2}. Therefore, the probability of the maximum being in any given $D \in \mathcal{C}(r)$ is bounded by $K s^2/n^2$, where K is a universal constant. Thus, in the terminology of Theorem 4.3,

$$\rho(r) \le \frac{K s^2}{n^2} = K' e^{-2r/C},$$

where $K' = K e^2$. We can assume that $K' > 1$. Again, if $x \in \mathbb{T}_n$ and $D \in \mathcal{C}(r)$ contains x, then the center of D must be at distance $\le 2s$ from x. Now, the centers of the members of $\mathcal{C}(r)$ are separated by distance $\ge s$ from each other. Clearly, the maximum number of s-separated points in a $2s$-ball can be bounded by a universal constant that does not depend on n and s. It follows that the number of members of D that contain x is bounded by a universal constant. Therefore, in the notation of Theorem 4.1, $\mu(r)$ is bounded by a universal constant. Using the bounds on $\rho(r)$ and $\mu(r)$ in Theorem 4.3 and the inequality $1 - x \le -\log x$ for $x > 0$, we see that for some universal constant κ,

$$\mathrm{Var}(\max_x \phi_x) \le C \log K' + \int_{C \log K'}^{C \log n} \frac{\kappa}{(2r/C) - \log K'} \, dr.$$

It is easy to see that the right hand side is bounded by a constant multiple of $\log \log n$. This completes the proof. $\square$

11. Example: Gaussian fields on Euclidean spaces

Consider a stationary centered Gaussian process $(X_n)_{n \ge 0}$. If we have $\mathrm{Cov}(X_0, X_n)$ decaying to zero faster than $1/\log n$ as $n \to \infty$, then it is known at least since the work of Berman [8] that $M_n := \max\{X_0, \dots, X_n\}$ has fluctuations of order $(\log n)^{-1/2}$ and upon proper centering and scaling, converges to the Gumbel distribution in the limit. This result has seen considerable generalizations for one dimensional Gaussian processes, both in discrete and continuous time. Some examples are [60], [61], [62], and [54]. For a survey of the classical literature, we refer to the book [46].

The question is considerably harder in dimensions higher than one. A large number of sophisticated results and techniques for analyzing the behavior of the maxima of higher dimensional *smooth* Gaussian fields are now known;

see the excellent recent book of Adler and Taylor [1] for a survey. Here 'smooth' usually means twice continuously differentiable. However, in the absence of smoothness, maxima of high dimensional Gaussian fields are still quite intractable. If only the expected size of the maximum is of interest, advanced techniques exist (see the book on chaining by Talagrand [75]). The question of fluctuations is much more difficult. In fact, for general (non-smooth) processes, even the one-dimensional work of Pickands [60, 61, 62] is considerably nontrivial.

One basic question one may ask is the following: what is a sufficient condition for the fluctuation of the maximum in a box of side length T to decrease like $(\log T)^{-1/2}$? In other words, when does the maximum behave as if it were the maximum of a collection of i.i.d. Gaussians, one per each unit area in the box? Classical theory [54] tells that this is true for stationary one-dimensional Gaussian processes whenever the correlation between X_0 and X_T decreases at least as fast as $1/\log T$. Note that this requirement for the rate of decay is rather mild, considering that it ensures that the maximum behaves just like the maximum of independent variables.

In the dimensions ≥ 2, the above question is unresolved. Here questions may also arise about maxima over subsets that are not necessarily boxes. Moreover, what if the correlation decays slower than $1/\log T$? In this section, we attempt to answer these questions. Our achievements are modest: we only have upper bounds on the variances. The issue of limiting distributions is not solved here.

Let $\mathbf{X} = (X_\mathbf{u})_{\mathbf{u} \in \mathbb{R}^d}$ be a centered Gaussian field on $\mathbb{R}^d$ with $\mathbb{E}(X_\mathbf{u}^2) = 1$ for each $\mathbf{u}$. For any Borel set $A \subseteq \mathbb{R}^d$, let

$$M(A) := \sup_{\mathbf{u} \in A} X_\mathbf{u}, \quad m(A) := \mathbb{E}(M(A)).$$

For any $\mathbf{u} \in \mathbb{R}^d$ and $r > 0$, let $B(\mathbf{u}, r)$ denote the open ball of radius r and center $\mathbf{u}$. Assume that

$$(46) \qquad\qquad L := \sup_{\mathbf{u} \in \mathbb{R}^d} m(B(\mathbf{u}, 1)) < \infty.$$

Note that in particular, the above condition is satisfied when the field is stationary and continuous. Next, suppose $\phi : [0, \infty) \to \mathbb{R}$ is a decreasing function such that for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$,

$$\mathrm{Cov}(X_\mathbf{u}, X_\mathbf{v}) \leq \phi(|\mathbf{u} - \mathbf{v}|),$$

where $|\mathbf{u} - \mathbf{v}|$ denotes the Euclidean distance between $\mathbf{u}$ and $\mathbf{v}$. Assume that

$$(47) \qquad\qquad \lim_{s \to \infty} \phi(s) = 0.$$

For a set $A \subseteq \mathbb{R}^d$ and $\epsilon > 0$, let $N(A, \epsilon)$ be the maximum number of points that can be found in A such that any two them are separated by distance greater than ϵ (such a collection is usually called an ϵ-net of A). When $\epsilon = 1$, we simply write $N(A)$ instead of $N(A, 1)$. The following theorem is the main result of this section.

Theorem 11.1. *Assume* (46) *and* (47). *Then for any Borel set* $A \subseteq \mathbb{R}^d$ *such that* $\text{diam}(A) > 1$, *we have*

$$\text{Var}(M(A)) \leq C_1(\phi, d)\left(\phi(N(A)^{C_2(\phi,d)}) + \frac{1}{\log N(A)}\right),$$

where $C_1(\phi, d)$ *and* $C_2(\phi, d)$ *are constants that depend only on the function* ϕ *and the dimension* d *(and not on the set* A*), and* $N(A)$ *is defined above.*

Remarks. The first observation is that if $\phi(s)$ decreases at least as fast as $1/\log s$, then the above result clearly shows that

$$\text{Var}(M(A)) \leq \frac{C(\phi, d)}{\log N(A)},$$

where $C(\phi, d)$ is some constant that depend only on ϕ and d. In particular, it gives a broad generalization of the classical results about fluctuations in one dimension [8, 60, 61, 62, 54]. An additional observation is that the first term in the bound can dominate the second only if $\phi(s)$ decreases slower than $1/\log s$.

Before we embark on the proof of Theorem 11.1, we need some simple upper and lower bounds on the expected value of $M(A)$.

Lemma 11.2. *Under* (46) *and* (47), *for any Borel set* $A \subseteq \mathbb{R}^d$ *such that* $N(A) \geq 2$, *we have*

$$c_1(\phi, d)\sqrt{\log N(A)} \leq m(A) \leq c_2(\phi, d)\sqrt{\log N(A)},$$

where $c_1(\phi, d)$ *and* $c_2(\phi, d)$ *are positive constants that depend only on the function* ϕ *and the dimension* d.

Proof. The upper bound follows easily from a combination of assumption (46), the tail bound from Proposition 1.3 for the maximum of the field in unit balls, and an argument similar to the proof of Lemma 2.1.

For the lower bound, first let $s > 1$ be a number such that $\phi(s) < 1/2$. Such an s exists by the assumption that $\lim_{s \to \infty} \phi(s) = 0$. Next, let B be a 1-net of A and D be an s-net of B. For each $\mathbf{x} \in D$, the s-ball around $\mathbf{x}$ can contain at most k points of B, where k is a fixed number that depends only on s and the dimension d. Thus,

$$|D| \geq \frac{|B|}{k} = \frac{N(A)}{k}.$$

Since $\phi(s) < 1/2$ and $\mathbb{E}(X_{\mathbf{u}}^2) = 1$ for each $\mathbf{u}$, by the Sudakov minoration technique from Section 2 we have

$$m(A) \geq m(D) \geq C\sqrt{\log |D|} \geq c_1(\phi, d)\sqrt{\log N(A)},$$

where D is a universal constant and $c_1(\phi, d)$ is a constant depending only on the function ϕ and the dimension d. □

Proof of Theorem 11.1. Let c_1 and c_2 be the constants from Lemma 11.2. Put

$$s := N(A)^{\frac{1}{8}(c_1/c_2)^2},$$

and assume that $N(A)$ is so large that $s > 2$. Let $r = \phi(s)$. Take any maximal s-net of A, and let $\mathcal{C}(r)$ be the set of $2s$-balls around the points in the net. It is easy to verify using the definition of s and the decreasing nature of ϕ, that $\mathcal{C}(r)$ is a covering of A satisfying the conditions of Theorem 4.3. Now take any $D \in \mathcal{C}(r)$. Since D is a $2s$-ball and $s > 2$, by Lemma 11.2 we have

$$m(D) \le c_2\sqrt{\log 2s} \le c_2\sqrt{2\log s} = \frac{c_1}{2}\sqrt{\log N(A)}.$$

Also, by Lemma 11.2, we have $m(A) \ge c_1\sqrt{\log N(A)}$. Thus, using the notation of Theorem 4.3, we have by Proposition 1.3 that

$$
\begin{aligned}
p(D) &\le \mathbb{P}\left(M(D) \ge m(D) + \frac{m(A) - m(D)}{2}\right) \\
&\quad + \mathbb{P}\left(M(A) \le m(A) - \frac{m(A) - m(D)}{2}\right) \\
&\le 2\exp\left(-\frac{(m(A) - m(D))^2}{8}\right) \\
&\le 2\exp\left(-\frac{c_1^2 \log N(A)}{32}\right).
\end{aligned}
$$

Since the above bound does not depend on D, it serves as a bound on $\rho(r)$. Let us now get a bound on $\mu(r)$. Take any $\mathbf{u} \in A$. Then the center of any $D \in \mathcal{C}(r)$ that contains $\mathbf{u}$ is a point in $B(\mathbf{u}, 2s)$. The centers are mutually separated by distance more than s. Hence, the number of $D \in \mathcal{C}(r)$ that can contain $\mathbf{u}$ is bounded by $N(B(\mathbf{u}, 2s), s)$, which, by scaling symmetry, is equal to $N(B(\mathbf{0}, 2), 1)$. Since $\mu(r)$ is the expected number of elements of $\mathcal{C}(r)$ that contain the maximizer of $\mathbf{X}$ in A, therefore we have that $\mu(r) \le c_3$, where c_3 is a constant that depends only on the dimension d. Combining the bounds, we see that whenever $N(A) \ge c_4$, we have

$$\frac{\mu(r)}{|\log \rho(r)|} \le \frac{c_5}{\log N(A)},$$

where c_4 and c_5 are constants depending only on ϕ and d. Note that we have this bound only for one specific value of r defined above. Now, if $r' > r$, and we define $\mathcal{C}(r')$ the same way as we defined $\mathcal{C}(r)$, then clearly $\mathcal{C}(r')$ would also be a cover of A satisfying the requirements of Theorem 4.3, and we would have $\rho(r') \le \rho(r)$. Noting this, and the fact that $|(1 - x)/\log x| \le 1$ for all $x \in (0, 1)$, we have by Theorem 4.3 that

$$\mathrm{Var}(M(A)) \le c_6\left(r + \frac{1}{\log N(A)}\right),$$

for some constant c_6 depending only on ϕ and d. Of course this holds only if $N(A) \ge c_4$, but this condition can now be dropped by increasing the value

of c_6, since we always have $\mathrm{Var}(M(A)) \leq 1$. Plugging in the value of r, the proof is done. $\qquad\square$

12. OPEN PROBLEMS

In some sense, this paper raises more questions than it solves. Some of these problems are listed below. At present, the author does not know how to solve any of these.

(1) Prove chaos in the original Sherrington-Kirkpatrick model, where $\xi(x) = x^2$. This is perhaps the most important open question that may be solvable by suitable extensions of the methods of this paper.

(2) Prove chaos in directed polymers in dimensions ≥ 3. Certain parts of the current proof do not work in higher than 2 dimensions, but other parts are okay. Since the Benjamini-Kalai-Schramm approach is inherently dimension-free, this seems to be a more doable open problem.

(3) Improve the fluctuation exponent in directed polymers. The current proof via hypercontractivity gives only a $\log n$ correction. Although this suffices to prove chaos, it is not satisfactory.

(4) Improve the variance bound in the discrete Gaussian free field with zero boundary condition. Although the case of the torus is interesting, the zero boundary condition is the more important one. Also, even in the torus, is $\log\log n$ the correct order? The author thinks $O(1)$ is more likely to be the right answer.

(5) Show that chaos implies multiple peaks even when the correlations are not all nonnegative. This involves getting a bound on the second moment of $R(I^0, I^t)$, which we do not know how to derive at present.

(6) Find a suitable multiple peaks condition that is actually equivalent to chaos and superconcentration. There is a possibility that there does not exist any such condition.

(7) Find a more intuitive and less analytical (rigorous) proof of the equivalence of superconcentration and chaos.

(8) A very interesting question in the context of multiple peaks is the issue of 'bridges': do the multiple peaks exist as disconnected islands, or do there exist 'bridges' that allow one to move from one peak to another without 'climbing down'? This question is particularly relevant for the Kauffman-Levin fitness model, for obvious evolutionary implications. Our method of proving the existence of multiple peaks suggests an obvious way to prove the existence of bridges, but the author does not yet know how to carry out the program for the NK model.

Nourdin, Jean-Philippe Bouchaud, Rongfeng Sun, and Giorgio Parisi for a number of useful discussions and communications. Special thanks go to Michel Ledoux for a major simplification of the proof of Theorem 3.2, to Rongfeng Sun for pointing out a serious mistake in the proof of Theorem 8.1 in the first draft (corrected by the author in the current manuscript), and to Itai Benjamini and Michel Talagrand for suggesting that the sensitivity to noise in glassy systems is a problem worth investigating.

References

[1] ADLER, R. J. and TAYLOR, J. E. (2007). *Random fields and geometry.* Springer Monographs in Mathematics. *Springer, New York.*

[2] AIZENMAN, M., LEBOWITZ, J. L., and RUELLE, D. (1987). Some rigorous results on the Sherrington-Kirkpatrick spin glass model. *Comm. Math. Phys.* **112** no. 1, 3–20.

[3] ALBEVERIO, S. and ZHOU, X. Y. (1996). A martingale approach to directed polymers in a random environment. *J. Theoret. Probab.* **9** no. 1, 171–189.

[4] ANÉ, C., BLACHÈRE, S., CHAFAÏ, D., FOUGÈRES, P., GENTIL, I., MALRIEU, F., ROBERTO, C. and SCHEFFER, G. (2000). Sur les inégalités de Sobolev logarithmiques. Panoramas et Synthèses, **10**. *Société Mathématique de France, Paris.*

[5] AUBRUN, G. (2005). A sharp small deviation inequality for the largest eigenvalue of a random matrix. *Séminaire de Probabilités XXXVIII,* 320–337, Lecture Notes in Math., **1857**, *Springer, Berlin.*

[6] BENJAMINI, I., KALAI, G., and SCHRAMM, O. (2001). Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math. No. 90 (1999),* 5–43.

[7] BENJAMINI, I., KALAI, G., and SCHRAMM, O. (2003). First passage percolation has sublinear distance variance. *Ann. Probab.* **31** no. 4, 1970–1978.

[8] BERMAN, S. M. (1964). Limit Theorems for the Maximum Term in Stationary Sequences. *Ann. Math. Statist.,* **35** no. 2, 502–516.

[9] BIGGINS, J. D. (1977). Chernoff's theorem in the branching random walk. *J. Appl. Probab.* **14** 630–636.

[10] BOLTHAUSEN, E. (1989). A note on the diffusion of directed polymers in a random environment. *Comm. Math. Phys.* **123** no. 4, 529–534.

[11] BOLTHAUSEN, E., DEUSCHEL, J.-D., and GIACOMIN, G. (2001). Entropic repulsion and the maximum of the two-dimensional harmonic crystal. *Ann. Probab.* **29** no. 4, 1670–1692.

[12] BORELL, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Invent. Math.,* **30** 205–216.

[13] BRAY, A. J. and MOORE, M. A. (1987). Chaotic Nature of the Spin-Glass Phase. *Phys. Rev. Lett.,* **58** no. 1, 57–60.

[14] CARMONA, P. and HU, Y. (2002). On the partition function of a directed polymer in a Gaussian random environment. *Probab. Theory Related Fields* **124** no. 3, 431–457.

[15] CARMONA, P. and HU, Y. (2004). Fluctuation exponents and large deviations for directed polymers in a random environment. *Stochastic Process. Appl.* **112** no. 2, 285–308.

[16] CATOR, E. and GROENEBOOM, P. (2006). Second class particles and cube root asymptotics for Hammersley's process. *Ann. Probab.* **34** no. 4, 1273–1295.

[17] CHEN, L. H. Y. (1982). An inequality for the multivariate normal distribution. *J. Multivariate Anal.* **12** 306–315.

[18] CHERNOFF, H. (1981). A note on an inequality involving the normal distribution. *Ann. Probab.* **9** 533–535.

[19] COMETS, F. and NEVEU, J. (1995). The Sherrington-Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case. *Comm. Math. Phys.* **166** no. 3, 549–564.

[20] COMETS, F., SHIGA, T., and YOSHIDA, N. (2004). Probabilistic analysis of directed polymers in a random environment: a review. *Stochastic analysis on large scale interacting systems,* 115–142, *Adv. Stud. Pure Math.,* **39** Math. Soc. Japan, Tokyo.

[21] COMETS, F. and YOSHIDA, N. (2006). Directed polymers in random environment are diffusive at weak disorder. *Ann. Probab.* **34** no. 5, 1746–1770.

[22] CRISANTI, A., PALADIN, G., SOMMERS, H.-J. and VULPIANI, A. (1992). Replica trick and fluctuations in disordered systems. *J. Phys. I France* **2** 1325–1332.

[23] DURRETT, R. and LIMIC, V. (2003). Rigorous results for the *NK* model. *Ann. Probab.* **31** no. 4, 1713–1753.

[24] EVANS, S. N. and STEINSALTZ, D. (2002). Estimating some features of *NK* fitness landscapes. *Ann. Appl. Probab.* **12** no. 4, 1299–1321.

[25] FISHER, D. S. and HUSE, D. A. (1991). Directed paths in a random potential. *Phys. Rev. B,* **43** no. 13, 10728–10742.

[26] FRÖHLICH, J. and ZEGARLIŃSKI, B. (1987). Some comments on the Sherrington-Kirkpatrick model of spin glasses. *Comm. Math. Phys.* **112** no. 4, 553–566.

[27] GARDNER, E. (1985). Spin glasses with *p*-spin interactions. *Nuclear Phys. B* **257** no. 6, 747–765.

[28] GIACOMIN, G. (2000). Anharmonic lattices, random walks and random interfaces. Recent Research Developments in Statistical Physics, Transworld Research, **I**, 97–118.

[29] GORDON, R. D. (1941). Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *Ann. Math. Statist.* **12** 364–366.

[30] GRAVNER, J., TRACY, C. A., and WIDOM, H. (2001). Limit theorems for height fluctuations in a class of discrete space and time growth models. *J. Statist. Phys.* **102** no. 5-6, 1085–1132.

[31] GROSS, L. (1975). Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** no. 4, 1061–1083.

[32] GROSS, D. and MÉZARD, M. (1984). The simplest spin glass. *Nuclear Phys. B* **240** 431–452.

[33] GUERRA, F. and TONINELLI, F. L. (2002). The thermodynamic limit in mean field spin glass models. *Comm. Math. Phys.* **230** 71–79.

[34] GUERRA, F. (2003). Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.* **233** 1–12.

[35] GUIONNET, A. and ZEGARLINSKI, B. (2003). Lectures on logarithmic Sobolev inequalities. *Séminaire de Probabilités, XXXVI,* 1–134, Lecture Notes in Math., **1801**, *Springer, Berlin.*

[36] HOUDRÉ, C. (1995). Some applications of covariance identities and inequalities to functions of multivariate normal variables. *J. Amer. Statist. Assoc.* **90** no. 431, 965–968.

[37] HOUDRÉ, C. and KAGAN, A. (1995). Variance inequalities for functions of Gaussian variables. *J. Theoret. Probab.* **8** 23–30.

[38] HUSE, D. A., HENLEY, C. L. and FISHER, D. S. (1985). Huse, Henley and Fisher respond. *Phys. Rev. Lett.* **55** no. 26, 2924.

[39] HWA, T. and FISHER, D. S. (1994). Anomalous fluctuations of directed polymers in random media. *Phys. Rev. B,* **49** no. 5, 3136–3155.

[40] IMBRIE, J. Z. and SPENCER, T. (1988). Diffusion of directed polymers in a random environment. *J. Statist. Phys.* **52** no. 3–4, 609–626.

[41] JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** no. 2, 437–476.

[42] JOHANSSON, K. (2001). Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. Math. (2),* **153** no. 1, 259–296.

[43] KAUFFMAN, S. A. and LEVIN, S. A. (1987). Towards a general theory of adaptive walks on rugged landscapes. *J. Theoret. Biol.* **128** 11–45.

[44] KIFER, Y. (1997). The Burgers equation with a random force and a general model for directed polymers in random environments. *Probab. Theory Related Fields* **108** no. 1, 29–65.

[45] KONDOR, I. (1983). Parisi's mean-field solution for spin glasses as an analytic continuation in the replica number. *J. Phys. A,* **16** L127–L131.

[46] LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. *Springer-Verlag, New York - Berlin.*

[47] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon.* Amer. Math. Soc., Providence, RI.

[48] LEDOUX, M. (2003). A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices. *Séminaire de Probabilités XXXVII,* 360–369, Lecture Notes in Math., **1832**,*Springer, Berlin.*

[49] LIMIC, V. and PEMANTLE, R. (2004). More rigorous results on the Kauffman-Levin model of evolution. *Ann. Probab.* **32** no. 3A, 2149–2178.

[50] MCKAY, S. R., BERGER, A. N., and KIRKPATRICK, S. (1982). Spin-Glass Behavior in Frustrated Ising Models with Chaotic Renormalization-Group Trajectories. *Phys. Rev. Lett.* **48**, 767–770.

[51] MEHTA, M. L. (1991). *Random Matrices.* 2nd ed. Academic Press, Boston, MA.

[52] MÉZARD, M. (1990). On the glassy nature of random directed polymers in two dimensions. *J. Phys. France* **51** 1831–1846.

[53] MÉZARD, M., PARISI, G., and VIRASORO, M. A. (1987). *Spin glass theory and beyond.* World Scientific Lecture Notes in Physics, **9**. *World Scientific Publishing Co., Inc., Teaneck, NJ.*

[54] MITTAL, Y. and YLVISAKER, D. (1975). Limit distributions for the maxima of stationary Gaussian processes. *Stoch. Proc. Appl.* **3** 1–18.

[55] NASH, J. (1958). Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.* **80** no. 4, 931–954.

[56] NELSON, E. (1966). A quartic interaction in two dimensions. *Mathematical Theory of Elementary Particles (Proc. Conf., Dedham, Mass., 1965),* M.I.T. Press, Cambridge, Mass., 69–73.

[57] NELSON, E. (1973). The free Markoff field. *J. Functional Analysis* **12** 211–227.

[58] NOURDIN, I. and VIENS, F. G. (2008). Density estimates and concentration inequalities with Malliavin calculus. Available at `http://arxiv.org/pdf/0808.2088v2`

[59] PANCHENKO, D. and TALAGRAND, M. (2007). On the overlap in the multiple spherical SK models. *Ann. Probab.* **35** no. 6, 2321–2355.

[60] PICKANDS, J., III (1967). Maxima of stationary Gaussian processes. *Z. Wahrs. und Verw. Gebiete* **7** 190–223.

[61] PICKANDS, J., III (1969). Upcrossing probabilities for stationary Gaussian processes. *Trans. Amer. Math. Soc.* **145** 51–73.

[62] PICKANDS, J., III (1969). Asymptotic properties of the maximum in a stationary Gaussian process. *Trans. Amer. Math. Soc.* **145** 75–86.

[63] SHCHERBINA, M. V. (1997). On the replica-symmetric solution for the Sherrington-Kirkpatrick model. *Helv. Phys. Acta* **70** 838–853.

[64] SHEFFIELD, S. (2007). Gaussian free fields for mathematicians. *Probab. Theory Related Fields* **139** no. 3-4, 521–541.

[65] SHERRINGTON, D. and KIRKPATRICK, S. (1975). Solvable model of a spin glass. *Phys. Rev. Lett.* **35** 1792–1796.

[66] DA SILVEIRA, R. A. and BOUCHAUD, J.-P. (2004). Temperature and Disorder Chaos in Low Dimensional Directed Paths. *Phys. Rev. Lett.,* **93** no. 1, 015901.

[67] SINAI, Y. G. (1995). A remark concerning random walks with random potentials. *Fund. Math.* **147** no. 2, 173–180.

[68] SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* **41** 463–501.

[69] SONG, R. and ZHOU, X. Y. (1996). A remark on diffusion on directed polymers in random environment. *J. Statist. Phys.* **85** no. 1–2, 277–289.

[70] SUDAKOV, V. N. and CIRELSON [TSIRELSON], B. S. (1974). Extremal properties of half-spaces for spherically invariant measures. (Russian) Problems in the theory of probability distributions, II, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **41**, 14–24, 165.

[71] TALAGRAND, M. (1994). On Russo's approximate zero-one law. *Ann. Probab.* **22** 1576–1587.

[72] TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* **81** 73–205.

[73] TALAGRAND, M. (1998). The Sherrington-Kirkpatrick model: a challenge to mathematicians. *Probab. Theory Relat. Fields* **110** 109–176.

[74] TALAGRAND, M. (2003). *Spin glasses: a challenge for mathematicians. Cavity and mean field models.* Springer-Verlag, Berlin.

[75] TALAGRAND, M. (2005). *The generic chaining. Upper and lower bounds of stochastic processes.* Springer-Verlag, Berlin, 2005.

[76] TALAGRAND, M. (2006). The Parisi formula. *Ann. Math. (2)* **163** no 1, 221–263.

[77] TINDEL, S. (2005). On the stochastic calculus method for spins systems. *Ann. Probab.* **33** no. 2, 561–581.

[78] TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** no. 1, 151–174.

[79] TRACY, C. A. and WIDOM, H. (1994). Fredholm determinants, differential equations and matrix models. *Comm. Math. Phys.* **163** no. 1, 33–72.

[80] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** no. 3, 727–754.

[81] TSIRELSON, B. S., IBRAGIMOV, I. A., and SUDAKOV, V. N. (1976). Norms of Gaussian sample functions. *Proceedings of the Third Japan-USSR Symposium on Probability Theory (Tashkent, 1975),* **550**, 20–41.

[82] WEHR, J. and AIZENMAN, M. (1990). Fluctuations of extensive functions of quenched random couplings. *J. Statist. Phys.* **60** no. 3-4, 287–306.

[83] ZHANG, Y.-C. (1987). Ground state instability of a random system. *Phys. Rev. Lett.,* **59** no. 19, 2125–2128.

367 EVANS HALL #3860
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA AT BERKELEY
BERKELEY, CA 94720-3860
E-mail: sourav@stat.berkeley.edu
URL: http://www.stat.berkeley.edu/∼sourav